

BJÖRN LINDBLOM

The goal of phonetics, its unification and application  
(Summary)

The full text will be published in the Proceedings of the congress,  
volume III, and in Language and Speech, 1980.

THE GOAL OF PHONETICS, ITS UNIFICATION AND APPLICATION

Björn Lindblom, Institute of Linguistics, 106 91 Stockholm,  
Sweden

When we compare the phonetics of today with that of the past we see progress. Looking ahead some of us may envision a glorious future for our discipline, others stagnation or even crisis.

Present-day phonetics differs in several ways from that of nineteenth century pioneers such as Passy, Sweet, Rousselot and others. We can point to the technological sophistication of our computers, speech synthesis or other experimental equipment, the development of an acoustic theory of speech or to the practical use that our understanding of human speech might be put to in various technological, educational and medical applications. It is also instructive to contrast past and present by recalling how classical phonetics dealt with the still current, fundamental problem of finding a universal phonetic framework for spoken language. This task is essentially that of describing phonetically an arbitrary utterance in any language (analysis) and to represent it in such a way that the description can be reproduced in audible form (synthesis) and with the linguistically relevant features (the original native accent) preserved.

The solution of classical auditory phonetics was the concept of the universal phonetic alphabet and the use of skilled phoneticians for the "recording" and "playback" of phonetic facts. However, this proposal fails. Its inadequacies cannot be remedied by invoking the insights contributed later by functional phonemic analysis and distinctive feature theory to define the terms "alphabet" and "universal" more precisely. Nor would it matter if the quest for the ultimate phonetic framework could be brought to a successful close and if suddenly phoneticians became capable of using it ideally. Contemporary phonetics rejects this solution since the scientific description of speech sounds must necessarily aim at characterizing explicitly and quantitatively - rather than merely skillfully imitating - the acoustic events as well as the psychological and physiological processes that speakers and listeners use in generating and interpreting utterances. Phoneticians accordingly construe their task of speech sound specification as a physiologically and psychologically realistic

modeling of the entire chain of speech behavior.

Experimental and theoretical progress up to now thus makes it possible to embed phonetics within a much broader intellectual context than previously. We might reasonably expect it to enjoy a favored position in future research on the forms and uses of spoken language in acquisition, production and perception. After all, why should it not be possible, on a long-term basis at least, for phoneticians to extend their inquiry into the sounds of human speech to ever deeper physiological and psychological levels using the speech signal as a window to the brain and mind of the learner, talker and listener? Why should we not expect more complete, theoretical models and computer simulations to be proposed for speech production, speech understanding and speech development that match the present quantitative theory of speech acoustics in rigor and explanatory adequacy? There seems to be particularly good reason for such optimism in the area of language universals where phonetics in fact has a privileged position. Linguistic behavior presumably arises, both ontogenetically and phylogenetically, as the result of an interplay between the (communicative, cognitive, social) functions that language is to subserve, biological prerequisites (brain, nervous system, speech organs, ear, psychological mechanisms such as memory etc.) as well as environmental factors. Languages thus evolve the way they do because of the body, the mind and the linguistic environment. They are the way they are on account of the functions they serve and owing to the properties of both innate and acquired mechanisms of learning, production and perception. This view assigns a novel and important future role to phonetics whose contents appears capable of offering general linguistic theory a great deal of explanatory force - a novel role at least to those who assign one major responsibility to phonetics in linguistics *viz.*, the instrumental analysis of the phenomena below the level of narrow phonetic transcription in grammars.

Looking back and ahead we see phonetics transform from more or less an art into a natural science. This development has yet to be completed but it is no doubt an inevitable consequence of the very nature of the subject matter of phonetics and the natural ambition of any discipline to attain scientific maturity. This trend has been and no doubt will be further stimulated by the

prospect of applying phonetic theory to practical needs such as pedagogical methods and technical aids for the deaf, handicapped, second language learners, the diagnosis and treatment of patients with phonetic symptoms as well as the automatic analysis and synthesis of speech for various technological purposes.

It may of course be objected that the program suggested above is entirely premature and unrealistic. It might be argued that, although it may be true that phonetics both could and should be pursued along such lines the practical difficulties must not be underestimated. At present it is far from a unified field. Progress so far seems often to have occurred in the form of fortuitous secondary spin-off effects from other adjacent fields with different goals rather than as a result of premeditated planning on the part of phoneticians and linguists. And by the way who is a phonetician these days? The heterogeneity of educational backgrounds in our field is striking. Recruiting researchers across disciplines has demonstrably had an extremely vitalizing influence. However, to meet the future challenge of developing a more comprehensive, unified phonetic theory will such heterogeneity be satisfactory? Will scientists coming into phonetics as basically faculty of arts students have the adequate training in mathematics and physics? Conversely will people trained in science and medicine have a chance to acquire the necessary background in linguistics and psychology and so forth? Who could claim the breadth and depth of competence that the present goals seem to imply? Perhaps we should accept that inevitably both applied and theoretical progress in our field has to occur on a basis of "mutual consultation" among a diversity of specialists. Science is a machine that develops very slowly under the influence of many forces and possibly more according to an open-loop mechanism than under the constraint of foresight and negative feedback. The problem boils down to that of adjusting research goals to the competence of the researchers or of adjusting the competence of the researchers to the research program. The former occurs easily enough. The latter requires more effort.

Although the preceding considerations are relevant and may serve to temper the optimism expressed earlier we shall conclude this summary on a positive note. Clearly there are active steps that can and should be taken to achieve a match between the



## 6 PLENARY LECTURE

training for a research career in applied or theoretical phonetics and the long- and short-term objectives of the field. There are also ways of achieving a greater unification of phonetics and eventually it is the questions asked that determine the future of a discipline.

## S p e e c h P r o d u c t i o n

Report: PETER F. MACNEILAGE  
 Co-Report: PETER LADEFOGED  
 Articulatory parameters  
 Co-Report: MASAYUKI SAWASHIMA  
 A supplementary report on speech production

11

41

49

## SPEECH PRODUCTION

Peter F. MacNeilage, University of Texas at Austin, Austin, Texas, USA

To borrow and adapt a phrase from the German psychologist Ebbinghaus (cited by Boring, 1950), the study of speech production has a long past but a short history. Interest in the speech production process was well developed as early as the time of Panini (Allen, 1953). The well established discipline of Articulatory Phonetics has been ostensibly, solely concerned with the production of speech, even though, according to G.O. Russell (cited by Ladefoged, 1975), phoneticians have been "thinking in terms of acoustic fact and using physiological fantasy to express the idea." But it has only been in the last 20-25 years that the scientific study of speech production, using sophisticated instrumental techniques, and the experimental method, has gained any momentum. Now, in 1979, we look at a flourishing discipline. My task is to convey the flavor of this discipline to a wide range of readers. Unfortunately, my own linguistic limitations prevent an adequate coverage of work not written in the English language. Different limitations dictate a neglect of various subtopics, hopefully to be corrected by my co-reporters. These subtopics include supra-segmentals, tone, timing, phonetic influences on sound patterns of languages, and many aspects of speech pathology. The status report falls into three sections: 1. Functional properties of the speech production apparatus; 2. Control principles underlying speech production; and 3. The biological basis of the speech production process.

1. Functional Propertiesa. Respiratory Function

The main function of the respiratory system during speech is to provide a relatively constant level of subglottal pressure which serves as the power source for the speech act. In long stretches of speech following deep inspiration, this is achieved by active muscular forces first combating and then complementing passive forces towards expiration (relaxation pressure) when lung volume is larger than its resting levels, and then combating passive forces towards inspiration when lung volume is smaller than its resting level. For many years now the work of Ladefoged and his

The reports will also be published in *Language and Speech*, 1980.

colleagues has provided the standard view of this process (e.g. Ladefoged, 1967). According to this view the following sequence of muscular events occurs: first inspiration is accomplished by the combined action of the diaphragm and the external intercostal muscles. In the initial stage of expiration, the external intercostal muscles combat the (expiratory) relaxation pressure. At the point when relaxation pressure alone becomes insufficient to maintain the required subglottal pressure, the internal intercostals begin to exert a gradually increasing expiratory effect. When lung volume becomes less than that at the end of normal expiration, other muscles including abdominal muscles begin to supplement the expiratory effects of the internal intercostals. Normal conversational speech involves a much more restricted range of lung volumes, and only the internal intercostal muscles are required for expiratory control.

Recent work by Hixon et al. (1976) appears to require one major modification of this view. These researchers consider the abdominal musculature to be continuously active under speech conditions, not only during expiration (for which they are anatomically suited) but during inspiration in conversational speech as well.

Hixon et al. consider that the role of abdominal muscle activity during expiration is to allow for more efficient alveolar pressure generation by the rib cage. This effect is explained by analogy with maneuvers that can be carried out with an elongated balloon in which the portion nearest the neck is analogous to the rib cage and the distal portion is analogous to the abdomen. If one squeezes the half of the balloon near the neck manually, to simulate rib cage maneuvers, pressure will build up within the balloon, and simultaneously cause the distal half of the balloon to expand outward. Combating this outward expansion by contracting the abdominal muscles allows a more efficient pressure build up immediately below the neck of the balloon when it is squeezed in that region.

The role of the abdominal muscles in inspiration is considered to be facilitation of the role of the diaphragm. It is noted that in comparison with quiet breathing, speech breathing consists of "extremely abrupt inspirations and considerably prolonged expirations and that short inspiratory periods are desirable for communication purposes". As Hixon et al. put it: "Because of

inward displacement of the abdominal wall, the diaphragm is displaced axially headward such that its principal muscular fibers (costal) become substantially elongated and its radius of curvature increased. The significance of this externally imposed adjustment is that the diaphragm is in effect "mechanically tuned" to a configuration that tends to optimize its potential for producing rapid and forceful inspiratory efforts."

In addition to providing a relatively constant subglottal pressure level for speech, the respiratory system provides transient increases in subglottal pressure for various suprasegmental and segmental purposes. The precise scope of this second role of the system is not yet well defined partly because of considerable methodological difficulties. Appropriate EMG data is hard to obtain. Body plethysmographs are limited in their sensitivity, and effects on subglottal pressure produced by changes in glottal resistance must be distinguished from effects due to activity of the respiratory system (Ohala, 1974).

#### b. Laryngeal Function

Laryngeal Mechanisms were dealt with quite comprehensively at the last International Congress (Fant and Scully, 1977) and there has not been a great deal of change in our knowledge about them since that time.

In the past few years there has been an increasing realization of the versatility of the vocal folds in producing vocal sound "at a wide range of fundamental frequencies, with great varieties of tonal qualities." (Hirano, 1977). The Myoelastic Aerodynamic Theory of van den Berg (1958) according to which the vocal folds are forced open by increasing subglottal pressure and close again as a result of their own elasticity and the Bernoulli force, remains an appropriate view of the phonation process. But in order to account for the wide range of conditions under which the vocal folds vibrate it has become useful to assume that each vocal fold consists not of a single mass but of a lower and an upper mass roughly corresponding to Hirano's dichotomy between the muscular "Body" of the folds and a mainly ligamentous "Cover", respectively. These two masses move to some extent independently during normal chest register phonation, partly because contraction of the vocalis muscle within each fold sufficiently counteracts the longitudinal tension effect of the cricothyroid to allow the cover to be "loose" and free to vibrate (Hirano, 1977).

We remain relatively uninformed as to how the vibratory patterns of the more unusual but nevertheless linguistically important modes of phonation such as a creaky voice and breathy voice are achieved. With respect to pitch control, a very straightforward relation seems to exist between pitch increase and activity of the cricothyroid muscle (e.g. Atkinson, 1978). On the other hand, there is yet little agreement on how pitch is lowered (Fujimura, 1977a). For one thing, the relative role of the passive effects of reduction in contraction of muscles associated with pitch raising, and the active effects of contraction of pitch lowering muscles situated extrinsic to the larynx has not been satisfactorily established.

Lisker and Abramson's (1971) contention that "the universally most important mechanism for the voiced-voiceless distinction is along the glottal adduction-abduction dimension" (Fant, 1977) is widely accepted. As Fant notes: "the posterior cricoarytenoid muscle... which is the only abductor would accordingly be responsible for glottal opening and thus devoicing in consonants irrespective of the degree of aspiration" (Fant, 1977). The interarytenoid muscle plays the main role in adduction.

The functional role of larynx height in the achievement of both the voiced-voiceless distinction, and the control of pitch is not yet understood. Larynx elevation is positively correlated with both devoicing, and pitch increase, but why this is the case has yet to be explained.

Of the three components of the speech production apparatus the laryngeal component has benefitted from the most sophisticated modelling of the interaction of aerodynamic and biomechanical influences. Prominent examples of recent models of vocal fold vibration are those of Flanagan et al. (1975) and Titze (1976).

#### c. Articulatory Function

The articulatory system is by far the most complex of the three components of the speech production apparatus. A great deal is now known about the way in which vocal tract area functions (shapes) serve to modulate the glottal sound source for speech. But in recent years our knowledge of vocal tract shapes has been pushed beyond the characterizations of traditional articulatory phonetics in two important ways. First, the postures actually adopted by the articulators have become better understood.

Second, using both traditional and new experimental techniques we have gained a good deal of new information about articulatory dynamics (Sawashima and Cooper, 1977). The following is a brief review of some examples of recent developments, intended to illustrate the diversity of motivation and method characteristic of this area of interest.

Some progress is signified by our greater readiness to accept the fact that speech is an "output oriented" activity (Fant, 1977, p. 8). Its aim is to produce an acoustic signal adequate to convey a linguistic message. Because of the non-uniqueness of the relation between vocal tract shapes and acoustic waveforms different speakers are able to communicate the same message with different articulatory postures. From the study of X-ray movies of 5 speakers, Ladefoged et al. (1972) showed that the traditional characterization of vowels in terms of the high-low and front-back dimensions of the tongue is not appropriate, and that there is considerable variation in the tongue configurations adopted by different speakers producing the same vowel. In a similar vein, Bell Berti (1975) has described individual differences in articulatory maneuvers assisting in control of the voiced-voiceless distinction, primarily by controlling vocal cavity volume so as to influence the pressure drop across the glottis.

The articulatory system consists of a set of interdependent structures innervated by a large number of muscles. Part of the search for functional principles underlying articulation has been an attempt to define the number of degrees of freedom in the operation of the system. For example Ladefoged and his colleagues (Harshman et al., 1977; Ladefoged, 1977) have used the statistical technique of Factor Analysis in an attempt to define the number of degrees of freedom in the production of tongue shapes for English vowels. Their analysis revealed two components, one representing "an upward and backward movement of the tongue", and the other representing "a forward movement of the tongue together with a raising of the front of the tongue" (Ladefoged, 1977, p. 217). Ladefoged notes that the former component can be thought of in terms of the action of the styloglossus muscle and the latter in terms of the action of the genioglossus muscle. But he cautions that the two components "if they have any physiological reality at all, are best thought of as high level cortical control functions." (p. 218).

A number of researchers have formulated articulatory models in an attempt to characterize various functional aspects of the articulatory system. Lindblom and Sundberg (1971a) have presented a model which is an attempt at an explicit quantitative specification of the contribution of the individual articulatory structures -- the lips, jaw, tongue and larynx (height) -- in the production of vowels. They also consider that tongue positions can be specified by 2 components, choosing the anterior-posterior location of the tongue body, and the extent to which the tongue body has been deformed from its natural shape. They justify the introduction of the jaw as a parameter in articulatory models on the grounds that it "makes it possible to explain why openness occurs as a universal phonetic feature of vowel production." In their view "the degree of opening of a vowel corresponds to a position of the jaw that is optimized in the sense that it cooperates with the tongue in producing the desired area function" (p. 1166).

An approach to modelling the physical properties of the tongue by computer simulation has been reported by Fujimura. The model "consists of 44 tetrahedral elements as internally uniform subunits of a linear elastic medium. These subunits are organized into 14 prism-shaped functional units representing independently controllable substructures." (Fujimura, 1977b, p. 226-7). The input forces, representing both intrinsic and extrinsic lingual muscles, "...can be specified as a linear combination of any number of internally uniformly distributed stresses within specified functional units, and forces acting directly (externally) on any of the nodal points of these units" (Fujimura and Kakita, 1978). The choice of forces is guided by EMG studies of the activity of lingual muscles during various speech gestures. An example of the outcome of this work, which is still in its earlier stages, is the interesting claim that the required vocal tract configuration for /i/ is relatively insensitive to the precise amount of contraction of the genioglossus muscle. This claim is analogous, at the physiological level, to the Quantal Theory of Stevens, based on observations of the relation between articulatory configurations and sound attributes: "For a particular range of an articulatory parameter, the acoustic output from the vocal tract seems to have a distinctive attribute that is significantly different from the acoustic attributes for some other region of the articulatory

parameter. Within this range of articulation, the acoustic attribute is relatively insensitive to perturbations in the position of the relevant articulatory structure." (Stevens and Perkell, 1977, p. 324).

If these approaches to articulation have anything in common, it is an attempt to define the constraints that determine the observed articulatory events and the absence of other articulatory events that seem logically possible. Attempts have also been made to specify constraints on articulatory dynamics associated with the production of stress and with changes in speaking rate. In an initial spectrographic approach to these questions Lindblom (1963, 1964) concluded that, in Swedish, vowel reduction in unstressed syllables and at faster speaking rates might simply be a mechanical result of the decreased time available for articulatory movements under these conditions. More recent EMG studies (e.g. Gay, 1977) have shown instead that there are differences in control signals to the articulators when stress level or speaking rate is changed. Apparently these stress and rate dependent changes in control cannot be accounted for in terms of any one simple algorithm. Consonants and vowels must be considered separately, as reduction effects are greater in vowels and segment durations reduce more in vowels than in consonants. Stress and rate effects are not always the same. Whereas vowel reduction is characteristic of unstressed syllables, it is only one of the 2 choices of an individual speaker in increasing speaking rate, the other being an increased rate of articulator movement to avoid reduction (MacNeilage, 1978a). Even reduction, when it occurs, is not simply accomplished by a uniform reduction in force of articulation. Amount of undershoot has been observed to differ on different vowels (Gay, 1977). The intuition that stress and rate modifications can be achieved by merely changing the values of some general time-dependent motor control variable has not yet been adequately supported.

## 2. Control Principles

Parallel with work on the functional properties of the speech apparatus has been a concern with the general control principles underlying speech production. Interest has focussed on easily identifiable articulatory gestures associated with individual speech segments -- particular tongue and lip configurations, jaw

positions and velar positions. A most deep-seated conviction is that there must be some invariance underlying the achievement of a configuration for a particular vowel or consonant, regardless of its segmental context. An early hypothesis was that this invariance might lie in the motor command sent to the muscles and observable by means of electromyograms (EMG). However, EMG studies showed, on the contrary, that context dependence in motor commands was the rule (MacNeilage, 1970). For other approaches I quote extensively from a recent review (MacNeilage, 1978a):

"Another group of theorists focussed on the fact that the results of gestures associated with a given phoneme (i.e. the positions achieved by them) remained relatively invariant in different contexts and suggested that therefore gestures were controlled in terms of the specification of invariant goals or targets. As to the nature of these goals or targets, I suggested in 1970 that they could be points specified within an internalized space co-ordinate system of the kind Lashley (1951) considered to underlie all movement control (MacNeilage, 1970). One indirect argument for this view is that visual-motor coordination is certainly guided by an abstract conception of space and therefore the auditory-motor coordination of speech may be also. In addition, control of the speech apparatus in the absence of an auditory component, as in the acts of mastication, and perception of oral stereognosis, would seem to require an abstract spatial analysis mechanism.

Informal evidence of the controlling role of goal or target specification during speech can be obtained by observing a speaker speaking with clenched teeth. Under this condition, acoustic output seems minimally impaired, suggesting that goals are successfully approximated, even though extensive compensatory articulation is probably required. More formal evidence comes from Lindblom and his colleagues who have twice performed an experiment in which subjects were required to produce vowels with bite blocks up to 25 mm in size between the teeth (Lindblom and Sundberg, 1971b; Lindblom et al., 1978). They found that immediately after bite block insertion, subjects achieved the correct formant frequencies in the first pitch period of the subsequent vowel. A subsequent midsagittal X-ray of these subjects during vowel production with a bite block inserted showed close approximation to normal vocal

tract shapes. This result suggests that even under the bite block condition articulators may be successfully controlled by invariant spatial goals or targets. However, in an experiment in which Folkins and Abbs (1975) unpredictably impeded jaw elevation movements associated with closure for a bilabial stop, the upper lip responded with active compensatory lowering, resulting in bilabial closure at a different (lower) point in space than normally observed. Such a finding suggests that the specification of goals or targets may not be in terms of absolute space in this case, but in terms of some other end such as articulator contact, or intra-oral pressure buildup. In addition, goals specified in terms of pressure would seem to be plausible in the respiratory system, where relatively constant subglottal pressure is preserved during speech, using widely varying muscular forces and lung volumes (Hixon et al., 1976)."

"In recent years a number of writers have emphasized the possible role of auditory targets in speech gesture control (e.g. Nootboom, 1970; Ladefoged et al., 1972). Informal evidence for the necessity of auditory targets in some sense of the term is quite conclusive. The auditory information provided by our language community is the only source of goals for our acquisition of speech production. A given auditory goal is sometimes achieved in a single subject by more than one spatial configuration of the speech apparatus. For example single intervocalic [p] is produced in English with vocal fold abduction (Lisker et al., 1969). But cluster-initial intervocalic [p] as in "upbringing" is produced in some subjects by vocal fold adduction (glottal stop) (Westbury, 1978). Thus it is the auditory goal that remains invariant in this case at the expense of invariance in spatial configurations. Further evidence on the relation between internalized auditory standards and movement control comes from an experiment by Riordan (1977). She reported that if rounding gestures of the lips are mechanically prevented, compensatory larynx lowering occurs, to achieve the lengthening of the vocal tract necessary to produce the formant frequencies of rounded vowels. This result shows that the control mechanism is capable of going beyond shape constancy in achieving auditory constancy.

The kinds of targets discussed so far are static targets. But when I produce the diphthong /au/ there is no evidence that

any static auditory or spatial target is being aimed at. In the period during which formant frequencies are relatively unaffected by preceding and following segments, the second formant for /au/ is in continuous motion. The perceptual importance of the dynamic properties of formant transitions, even for vowels, leads us to believe that some specification of dynamic properties must underlie the talker's production of them. Of course, close specification of the dynamics of speech movements is always made by the talker in an utterance whether it has any obvious perceptual consequences or not. Thus the issue to be raised here, ----, is the relation between static and dynamic aspects of the operation of the system." (MacNeilage, 1978a).

A good deal of work has been done on coarticulation, the study of the temporal scope of particular articulatory gestures and how this changes with segmental context. Coarticulation effects have been of interest because of the hope that the precise temporal scope of these effects would provide us with an understanding of the role of various linguistically defined units (e.g. the phoneme, the distinctive feature, the syllable, the word) in the movement control stages of the speech production process. Coarticulatory effects have been observed for up to 7 segments in the anticipatory (right to left) direction (Benguerel and Cowan, 1974) and in the perseveratory (left-to-right) direction (Ghazelli, 1977). Although they occasionally seem to respect syllable boundaries (Ushijima and Hirose, 1974) and word boundaries (Ghazelli, 1977), more often their temporal scope seems independent of the boundaries of linguistic units. They are sometimes not even blocked by mechanical incompatibility between the coarticulatory gesture and gestures for other segments (Sussman et al., 1973). The only thing that seems to reliably block these effects is the avoidance of production of an "immediate successional impact" -- a change in the acoustic properties of a neighboring segment which would change its message status for the listener (Kent and Minifie, 1977). Thus, all in all, the use of coarticulation to determine the basic properties of the control system has been relatively unsuccessful.

In conclusion, it must be conceded that we still know very little about the issue of invariance in the control of gestures or about the principles underlying coarticulation. What we have done so far is little more than to point to aspects or consequences

of gestures that possess invariance and suggest that the goal of the control system must be to achieve this invariance.

In some sense, what we are seeking is biological equivalents of linguistic units. But the precise relation between linguistic units defined primarily by means of analysis of the message structure of language, and control units, compatible with speech signal characteristics, is extremely hard to define. As Stevens and Perkell (1977) point out "There is little argument among students of speech and language that speech events at one level are organized in terms of segments and features." (p. 323). But the lack of argument may only exist because there has been comparatively little effort to reconcile the message and signal levels of conceptualization. The two researchers who have made the greatest recent effort to characterize speech from the traditional viewpoint of articulatory phonetics, Catford (1977) and Ladefoged (1975, 1978) both warn against assuming any simple relation between signals and message units. Catford has concluded that the attempt to define a finite universally applicable set of distinctive features is at best procrustean. Ladefoged concludes that it is erroneous to assume that a phonological feature can be defined in terms of a single physical scale which can be used for specifying contrasts between and within languages. He argues that: "From an acoustic or physiological point of view most phonological features are cover features definable only in terms of complexes of phonetic parameters." In the absence of a straightforward biologically defined relation between observable signal properties and underlying message units it is not clear what it means to assert that speech events at one level are organized-- "in terms of features".

Speech errors have been an important source of the comparatively rare information which bears on what one can call the psychological reality of linguistic units (Fromkin, 1973). An underlying assumption has been that linguistic units have psychological reality to the extent that they show themselves, in speech errors, to be independently variable in the time domain. Distinctive features do not appear to qualify as psychologically real units on these grounds. In approximately two-thirds of spoonerisms involving segments the target segments (e.g. [l] and [r] in "leaf raking" "reaf laking") differ by only 1 distinctive feature



(MacKay, 1970) and thus it is not possible to decide whether a feature or a segment has been exchanged. But where target segments differ by more than one feature (e.g. [p] and [n] in "pointed nail") there is almost never an exchange of a single feature (e.g. "tainted mail") even though, in terms of a feature model, that would be the simplest error. Shattuck-Hufnagel and Klatt (1978) have recently concluded that "features are not independently movable entities at the level where most substitution and exchange errors are made." Neither do syllables move around in speech errors. Nor do speech error data encourage the choice of the phoneme as an underlying form at the segmental level because segmental permutations are so restricted by phonotactic factors -- prevocalic, vocalic, and post-vocalic segments exchange with like components -- whereas the phoneme is ideally a context-free entity. Perhaps the most basic unit at the level at which most temporal sequencing errors can occur is the phonotactically restricted allophone.

Speech errors have also been used to try to determine the number of relatively separate stages or levels in the speech production process and the operations which occur at those levels. It remains possible, though difficult to demonstrate that units such as the feature, phoneme, and syllable have psychological reality at levels where operations do not lend themselves to independent variation of those units in speech errors. (For example it has been suggested (Fry, 1964) that the syllable is a rhythmic entity.) We are still a long way from being able to place sufficient constraints on multistage schemata for speech production. Two illustrations of the problems for such schemata can be given by considering the relation between direct movement control and more underlying levels. On the one hand it has been observed, as mentioned earlier (Westbury, 1978) that the opposite motor control gestures of vocal fold abduction and adduction can be used to achieve examples of what is assumed to be the same underlying voiceless consonant. On the other hand similar motor control gestures can be used to achieve examples of opposing underlying forms. The timing of voice onset and closure release is similar for representation of underlying /d/ in "duck" and (presumably) underlying /t/ in "stuck".

A number of researchers have shown some impatience with efforts to determine the nature of speech control at levels so far

away from direct observation, at the present stage of our knowledge (e.g. Moll et al., 1977). The following paraphrase of the views of Netsell (Moll, 1977) is representative of the concerns of these researchers:

"What existing articulatory data allow us to differentiate the nature of the input commands as phonemes, phones, syllables, words, etc.? Given present methodologies and conceptualizations, what would be the nature of an experiment or experiments that would clarify the character of these motor commands? In relation to these questions, it was noted that inferences from articulatory data can result in varying and contradictory conclusions concerning the nature of the input commands. This results from the fact that articulatory measurements reflect both the effects of the input commands, presumably at a high neural level, and the properties of the intervening physiological systems. Thus, it will not be possible to make inferences about the input command structure until we can separate out the effects of the system characteristics. In addition, it was noted that we must be able to formulate our hypotheses about command units in unambiguous and physiological terms before we can test them effectively by physiological observation." (p. 407).

Netsell's viewpoint emphasizes the importance of understanding the peripheral neuromechanical properties of the speech production system, and from this perspective the relative role of feedback or closed loop control becomes a central issue. On this issue most would agree with Stevens' (1977) view that: "Since speech can be considered to be a habituated or stereotyped form of motor behavior (which does not usually encounter external disturbances), preplanning mechanisms are probably used much more heavily than peripheral feedback for the moment-to-moment control of vocal-tract movements." (p. 343-346). However, Abbs and his associates have pointed out that: "regardless of the nature of the underlying neuromuscular mechanism controlling speech production, evaluation of the system biomechanics is necessary" (Muller et al., 1977). They note that: "if an open loop mode of control is hypothesized then we must concede that the system is either quite knowledgeable about the biomechanics and compensates for them during the control process of organizing the movement, or that the biomechanical characteristics are relatively simple and need not



be considered by the central mechanism." Up until the last few years, most conceptions of the upper articulators assumed a third possibility. Mechanical properties were considered to limit a speaker's ability to produce an invariant output for a phonological unit, resulting in undershoot and coarticulatory effects which the communication system had to tolerate because of neural limitations in production control. (See MacNeilage, 1970, for a review of this view.) But as Abbs and Eilenberg (1976) observed: "it cannot be assumed that peripheral mechanical influences are limiting in their influence upon speech movements. The passive properties of inertia and elasticity most appropriately are considered energy storage mechanisms, and although they may absorb energy generated during one interval of time, they have the capability to release that energy for later contributions to the system's output." (p. 142).

It is now well known that the control system has developed the ability to take stored elastic energy into account in the respiratory system. This was evident in the account of respiratory control given earlier. But the relative role of open and closed loop control in the ongoing control of these and other aspects of speech production is not yet well understood. Part of the motivation behind the work of Abbs and his associates on this topic is the assumption that by systems analysis it can be determined whether it is even analytically possible for any articulator movement to be under closed loop control given the transfer function of the particular neuromechanical system under observation, and the observed movement dynamics (Muller et al., 1977; Abbs et al., 1977). Variables involved in this transfer function include mechanical properties of the musculature and its accompanying load, sensory receptor properties, and potential neural transmission circuits with their loop gains. Mechanical and transmission delays in any system are detrimental, because they create phase lags between the system error and the feedback correction signal. Abbs et al. (1977) state that "... it is possible to determine the bandwidth over which a particular afferent loop might contribute; depending on the total phase lag introduced by its components. That is, if a feedback loop is to contribute to the control of movement, it must have a positive gain in the same frequency range as the movement itself."

In recent years there has been a good deal of disillusionment with the main technique used to investigate closed loop con-

trol of speech -- the sensory nerve block technique -- because of the lack of specificity inherent in its application and the possibility that it involves direct motor effects (Borden et al., 1973). Using the more acceptable technique of on-line intervention, Folkins and Abbs have unequivocally shown the operation of closed loop control of jaw elevation by demonstrating a compensatory response to interference with elevation which had a latency of 30 ms. But as Abbs et al. have pointed out, a general answer to the question of the relative role of closed loop control during speech must await a detailed analysis of all the individual system components in the entire speech apparatus.

Another approach to the understanding of peripheral speech mechanisms, favored in our laboratory, is the analysis of the final form of the neural control signals, at the level of the individual motor unit. The motor unit (an individual motoneuron and the muscle fibres that it innervates) has been described as the quantal element of movement control. The entire control signal for speech production can be described in terms of two variables of motor unit function, the number of motor units activated, and the discharge rate of each. We have concerned ourselves with the understanding of these two variables in speech musculature. In addition to providing parametric information which should aid in modelling of the neuromuscular stage of speech, we have determined that (with minor qualifications) the Size Principle (Henneman et al., 1965) is operative in speech control as in many other aspects of human and animal motor control (MacNeilage, 1978b). "Size" here refers to a number of correlated properties. For example: large motor units have larger cell bodies and larger axons, and their axons innervate more and larger muscle fibers. A number of functional properties are considered to relate to cell body size. Larger cell bodies are considered less excitable than small ones and are thus recruited into a movement at higher input levels (that is, later) than small ones. But once activated, larger cell bodies have greater sensitivity to input changes than small ones. They can be said to have higher gains. Larger cell bodies also have shorter afterhyperpolarization (AHP) durations than small ones. For our purposes AHP duration can be regarded as an index of a recovery cycle following a motoneuron discharge (a single firing) so that short AHP durations allow higher discharge rates. This information

can be regarded as a contribution towards defining the terms in which any control decisions are communicated to the speech apparatus.

As a footnote to this section on control of speech production, I give notice that an alternative view to conventional theories of speech production has recently been advanced. Again I quote extensively from a recent review (MacNeilage, 1978a): "The view arises from what has come to be known as Action Theory which has as its aim a general theory of coordinated movement. (Perhaps a better term at present would be Action Metatheory as it consists primarily of ideas about the form that theory of action should take.) Action Theory owes its origin primarily to the Russian physiologist Bernstein (1967) and has been developed in this country particularly by Greene (1972) and Turvey (1977) and with respect to speech by Fowler (1977, Fowler et al., 1978). The theory calls for a radical reformulation of the theory of speech production. Current speech production theories, that assume underlying units, and elaborate processes governing the surface manifestation of these units are dubbed "translation theories" (Fowler et al., 1978). These theories are considered to be unnecessary, as the linguistic units are, in some sense, directly and invariantly represented in the output, and do not exist independent of that representation. Temporal and spatial aspects of control are regarded as being integral to each other and therefore not to be considered separately. The two components of current models, the basic segmental specification, and timing schemes are thus also considered integral to each other and timing is described as intrinsic rather than extrinsic. This means that the timing-determined properties of the output arise naturally from its intrinsic organization.

The central concept of action theory is that of the "coordinative structure". A coordinative structure is defined as "a set of muscles, often spanning many joints, that is constrained to act as a unit" (Turvey et al., 1978). These structures are considered to be established by biasings of reflex circuits referred to as "tunings". Some properties of coordinative structures are modulable; "For example rate of walking and gait are modulable properties of the muscle systems that determine walking" (Fowler, 1977, p. 206). The constraints arising from the coordinative structures

are considered to determine directly what effects the modulable properties will have on movement. For example, I assume speaking rate would be a modulable property of the coordinative structures governing speech production, and thus its effects would arise directly from those structures rather than being imposed on them from an external source.

An act such as an utterance "is believed to be governed by functionally embedding (as opposed to temporally concatenating) coordinative structures. Each nesting level delimits a broader equivalence class of movements than the finer grain level nested within... The more coarse-grained nesting levels are established by altering the relationships among smaller coordinative structures, and at the same time they act as constraints on the lower ones." (Fowler et al., 1978, p. 28). For example, in Fowler's (1977) analysis the most coarse-grained coordinative structure embodies an entire utterance and at the other extreme 4 coordinative structures are proposed for vowels.

A good candidate for one of the most coarse-grained coordinative structures could be one that is responsible for maintaining a relatively constant level of subglottal pressure during a single expiratory phase. One coordinative structure for vowels is considered closely analogous to the state underlying the voluntary assumption of a certain fixed joint angle at the elbow, by a human subject, and the maintenance of that angle in the face of various loads (Asatryan and Fel'dman, 1965). The subject is considered to adopt an arbitrary "zero-state" of the joint-muscle system, thus creating a system with spring-mass properties. The production of /ε/ is considered to be achieved partly by the imposition of a "zero-state of the extrinsic tongue system" (Fowler et al., 1978, p. 71) which will produce tongue elevation following /æ/, but tongue depression following /i/.

It is obviously not possible to do justice to this new theoretical orientation in the space available here. The theory is provocative in its attempt to place speech production in a general biological perspective, and in its implication that speech production is by no means special. Nevertheless it is my overall impression that the relation between traditional theories of speech production and Action Theory has so far only been loosely defined, and that the value of the analogies made between speech and other

coordinated movement sequences still needs to be carefully scrutinized. The specific consequences of Action Theory for speech production have not yet been well established. For example, are changes in vowels with stress and speaking rate consistent in form with some specific coordinative structures with specific modulable properties? Is what is descriptively labelled as "undershoot", a result of changes in the coordinative structures or in the effects of modulable properties? How does the theory handle the case mentioned earlier of two opposite movement outcomes (abduction and adduction) in the achievement of voicelessness for /p/? Do short term memory constraints influence the operation of coordinative structures? It is to be hoped that these and many other problems will be fruitfully addressed as the implications of Action Theory for speech become clearer, and the interface between Action Theory and traditional ideas about speech production control becomes more clearly defined."

### 3. Biological Basis

Naturally occurring language -- including speech -- is still considered species-specific to humans despite the inroads being made by chimpanzees and gorillas learning sign language. But the biological basis for this specificity is not yet established. The considerable linguistic ability displayed by chimpanzees and gorillas using manual signs, together with the failure to teach them language involving speech production, suggests that species-specificity is greater for speech than for language (in the sense of an abstract message system). Species specificity in speech perception ability has not yet been established (though see Warren, 1976). The one-month-old human has shown a spectacular ability to discriminate between linguistically distinctive stimuli, most marked for the voice onset time (VOT) dimension (Eilers, 1978). But this ability, though apparently innate, cannot necessarily be considered either species specific or speech specific, in the light of the perceptual ability displayed by chinchillas with the VOT continuum, and the results of perceptual studies of nonspeech analogs of the VOT continuum (MacNeilage, 1977) including a study of infants (Jusczyk et al., 1977).

With respect to speech production, Lieberman (1975) has made a case that a "crucial" evolutionary development separating modern humans from lower hominids, is the development of the "bent two-

tube" supralaryngeal vocal tract from an earlier single tube configuration. This structural development increases the possible number of vocalic items in a speech sound inventory and makes possible production of the 3 "quantal" vowels /i/, /u/ and /a/. Lieberman (1972) has claimed that: "Modern speakers, in all likelihood make use of these extreme vowels to ascertain the size of the vocal tracts of individual speakers. This information is essential for the speech "decoding" that is the basis of the rapid rate of information transfer of human speech. Neanderthal man, though he could produce part of the human phonetic repertoire, would be incapable of speaking any human language" (p. 272). However, more recently, Verbrugge et al. (1976) have concluded that with respect to perception, "There is little evidence to support a claim of a special role for the point (quantal) vowels." (p. 198). Perhaps the structural development reported by Lieberman has only quantitative significance, making more distinctive sounds possible, though it should be noted that some human languages have quite small segmental inventories (e.g. 15 in New Zealand Maori (Biggs, 1961)).

On the topic of functional aspects of the biological basis of speech production, it is my contention that we have seriously underrated the importance of the "long and for the most part orderly series of stages of prelinguistic vocalization that are quite stereotyped in nature, and occur in the absence of model vocalizations of others" (MacNeilage, 1978a). A case can be made for the innateness and species specificity of many of these vocalizations including babbling (chimpanzees are considered to babble only in a quite restricted sense (Kortlandt, 1973)), and despite the methodological problems attendant on their study, they would seem to be a valuable source of information about the biological basis of speech. The neglect of prelinguistic vocalizations is, in my view, largely the result of the influence of Jakobson's theory of language acquisition (Jakobson, 1968), to which babbling was irrelevant. My view that prelinguistic vocalizations are important comes from recent studies by Oller et al. (1976) and others whom they cite, which show an extremely close relation between the sounds and sound sequences of babbling, even the earliest babbling (6-8 months) and the sounds and sequences used in the first words.

As to sounds, aspirated stops, fricatives and liquids are rare in babbling and in first words while unaspirated stops and glides are frequent in both. As to sequences (or phonotactic constraints), consonant clusters and final consonants (especially voiced final consonants) are rare in babbling and in first words, while initial consonants, especially stops, are common in both. Oller et al. conclude: "after examining our data on babbling it is possible to predict quite accurately the nature of the most commonly reported substitutions and deletions which occur in meaningful child speech." (1976, p. 9). I would go further and guess that it is possible to predict from the babbling of an individual child a great deal about what his/her first words will be like, including initial sounds and sequences, relative preference for reduplicated syllables, and aspects of temporal control such as segment durations and variability in those durations.

From this viewpoint the child's first words can be seen as, "at least partially, a matter of choosing from the babbling repertoire a set of approximations to adult word forms. As the babbling forms are quite limited, the first words represent an enormous simplification of adult forms, and subsequent learning can be seen primarily as the relaxation of constraints on earlier articulated forms as ability to produce additional forms increases. This view can be contrasted with Jakobson's view of first word learning as a matter of the unfolding of a fixed universal sequence of sound contrasts. (For a review of evidence against this view, see Ferguson, 1978.) The matter of choice of the child's first words was considered by Jakobson to be the result of "a selection by which they become speech sounds only insofar as they are related to language in the strict sense of the word. The selection is therefore inseparably linked to the sign nature of language, that is, is a purely linguistic matter." In contrast: "The question of the prelanguage babbling period proves to be, on the contrary, one of external phonetics, predominantly articulatory in nature ..." (1968, p. 27). A problem of Jakobson's viewpoint is that an appeal to the sign function of sounds gives no explanation of the particular order in which sounds do tend to appear in the first words. (The notion of maximal sound contrast has quite limited explanatory power in this respect.) But if one considers speech as an extension of babbling and it is seen that the first sounds of speech

tend to be the first sounds of babbling then a good deal of the order in sound acquisition becomes at least potentially explicable in terms of the biology of the production mechanism. One thus loses the need for innate components of speech such as a sign function. The sign function of language may evolve in the framework of the biology of the signalling mechanism and not vice versa. Two examples may illustrate this point. The voiceless unaspirated stop consonant is the only universal stop consonant category and it is the first stop consonant to be observed in babbling and in the child's first words. The Consonant-Vowel syllable is the only universal syllable type, and it is the first to be observed in canonical form (i.e. with the time-space constraints typical of speech) in the babbling stage and in the child's first words. Babbling can be regarded as the functional skeleton on which child speech is built and the functional skeleton on which sound patterns of languages are built.

There are two important qualifications to be made about the view expressed here. First, the tendency towards a fixed order of acquisition of sounds is just a tendency. Numerous individual exceptions have been noted. These exceptions mean that no theory that includes a single fixed order of acquisition of sounds or rules, can be valid. Second, although much emphasis has been placed here on prelinguistic vocalizations being carried over into speech without change in form, the functional significance of the babbling stage does not inhere in these forms alone. Other non-canonical forms such as the popular (universal?) 2-3 second long "raspberry" (Oller, 1978) may serve some purpose. In addition both my children (and others) have favored the production of C-V alternations with lateral manual stop consonants, produced by placing the back of the hand in a horizontal position over the mouth at the proximal finger joints, and alternately flexing and extending the fingers. These behaviors suggest that the babbling stage provides a functional skeleton for speech production in a sense that includes, but also transcends, motor stereotypies produced with the time-space properties and the apparatus later used for speech production.

In summary, I believe a case can be made for the innateness and species-specificity of many aspects of prelinguistic vocalization, which may provide a functional skeleton for the development

of the speech production process, both in terms of specific motor behavior and in terms of more general functional prerequisites.

Another source of evidence that has been widely used to claim an innate species-specific basis for language and speech is that of left hemisphere specialization. However, it is important to note that this evidence cannot be used to support a "speech is special" notion. Studies by Kimura and her associates (Kimura and Archibald, 1974; Mateer and Kimura, 1977) have shown that left hemisphere damage, whether accompanied by aphasia or not, results in impairment of reproduction of sequences of movements, whereas right hemisphere damage does not. These results have led Kimura (1976) to suggest that underlying left hemisphere specialization for language and speech may be specialization for the control of skilled motor sequences. In addition, recent studies show that chimpanzees have an anatomical differentiation between left and right hemispheres analogous to that found in humans (Galaburda et al., 1978). These findings, taken together with the failure to induce vocally based linguistic communication in chimpanzees, encourage us to view left hemispheric specialization for speech in man within a broad biological perspective. It is my belief that a good deal of our present knowledge is consistent with the following proposition: left hemisphere specialization for language and speech is derivative of the specialization of the left hemisphere for the control of skilled voluntary movement sequences, and for the perceptual analysis of stimuli related to the movement control skills.

As to the means by which the left hemisphere controls speech production, most speculation, based primarily on aphasia, has been directed towards the localization of functions in the cerebral cortex. Very little information is available about exactly what these functions are, and how they are controlled. The most acceptable general proposition is that whereas damage to anterior cortex results in a nonfluent aphasia, most typically classifiable as Broca's Aphasia, damage to posterior cortex results in a fluent aphasia typically classifiable as Wernicke's Aphasia. In the well known schema of Geschwind (1972) the speech production deficits of Wernicke's Aphasia are due to damage to the mechanism responsible for auditory patterns controlling output, and Broca's Aphasia deficits are in production of speech movements from the auditory

patterns. An additional syndrome of Conduction Aphasia, (now widely accepted by aphasiologists (Green and Howes, 1977) was considered due to a disconnection between the auditory and motor centers resulting from a lesion of the arcuate fasciculus, in the parietal lobe. In the light of the probable role of somesthetic functions, including spatial conceptualization in speech production control, and in light of the parietal lesion site for Conduction Aphasia, I have suggested (MacNeilage, 1978b) that Geschwind's explanation of Conduction Aphasia be re-examined and perhaps supplemented with another. In this view, Conduction Aphasia could at least in part be regarded as a deficit in somesthetic conceptualization underlying spatial target assignment for speech movement control. There were 4 reasons for this suggestion which are discussed in detail elsewhere (MacNeilage, 1978b): "First, the dominant speech production symptom of phonemic paraphasias is consistent with this view. Second, conduction aphasia is typically, though not always associated with apraxia, and apraxia would be expected to result from a deficit in spatial target function. Third, the lesions associated with the syndrome are in the parietal lobe which has traditionally been associated with spatial functions. Fourth, there is no alternative explanation in the literature that has a good claim to being preferable to the one I am suggesting."

Since this view was put forward, apparent counterevidence has come from studies of cerebral blood flow as an index of localization of cortical activity during speech production (Lassen et al., 1978). According to this index, posterior inferior parietal cortex activity was as low as that in several other sites, not expected to be involved in speech production (e.g. occipital cortex). As expected, high levels of activity were observed in Wernicke's Area and in primary sensorimotor cortex serving the speech apparatus (pre- and post-central gyri) and some elevation of activity was observed in Broca's Area. In addition, a surprising finding was the high levels of activity observed in a relatively large region of Supplementary Motor Cortex during speech as well as other motor functions. This region, in superior medial frontal cortex was implicated in speech control by the work of Penfield and his associates (Penfield and Roberts, 1959) but has received little attention since that time. Lassen and his associates con-

sider that the Supplementary Motor Cortex "is involved in the planning of sequential motor tasks" (p. 69). But although the findings of Penfield and Roberts corroborate the blood flow studies with respect to the involvement of Supplementary Motor Cortex (and Wernicke's and Broca's areas), they did not concur in observing a lack of involvement of posterior inferior parietal cortex. Evidence from both electrical stimulation and excision strongly implicated the parietal area in speech control. A resolution of this dilemma apparently awaits a better understanding of the meaning of the evidence from these 3 different sources.

Another finding of the work on cerebral blood flow was that of high levels of activity in the right hemisphere analogs of the left hemisphere regions which showed high activity during speech. The rarity of aphasia from right hemisphere lesions, the absence of right hemisphere initiation of speech in commissurotomy patients and the very limited scope of right hemisphere initiation of speech in left hemispherectomized patients (Searleman, 1977) suggests that right hemisphere does not play a necessary role in speech production. Nevertheless the blood flow studies suggest that the right hemisphere may normally share in speech production to a greater extent than had been previously supposed. The result also suggests that patients with left hemisphere lesions may be able to make more use of the right hemisphere to control speech than most current views of aphasia imply.

#### References

- Abbs, J.H. and G.R. Eilenberg (1976): "Peripheral mechanisms of speech motor control", in Contemporary issues in experimental phonetics, N.J. Lass (ed.), 139-168, New York: Academic Press.
- Abbs, J.H., E.M. Muller, M. Hassul, and R. Netsell (1977): "A systems analysis of possible afferent contributions to lip movement control", paper presented to the American Speech and Hearing Association, Chicago, 1977.
- Allen, W.S. (1953): Phonetics in ancient India, London: Oxford University Press.
- Asatryan, D. and A. Fel'dman (1965): "Functional tuning of the nervous system with control of movement or maintenance of a steady posture -- I. Mechanographic analyses of the work of the joint on execution of a postural task", Biophysics 10, 925-935.
- Atkinson, J.E. (1978): "Correlation analysis of the physiological factors controlling fundamental voice frequency", JASA 63, 211-222.
- Bell-Berti, F. (1975): "Control of pharyngeal cavity size for English voiced and voiceless stops", JASA 57, 456-61.
- Benguerel, A.-P., and H.A. Cowan (1974): "Coarticulation of upper lip protrusion in French", Phonetica 30, 41-55.
- Bernstein, N. (1967): The coordination and regulation of movements, London: Pergamon Press.
- Biggs, E. (1961): "The structure of New Zealand Maaori", Anthropological Linguistics 3, 1-53.
- Borden, G.J., K.S. Harris and L. Catena (1973): "Oral feedback II. An electromyographic study of speech under nerve block anesthesia", JPh 1, 297-308.
- Boring, E.G. (1950): A history of experimental psychology, New York: Appleton Century Crofts.
- Catford, J.C. (1977): Fundamental problems in phonetics, Bloomington: Indiana University Press.
- Eilers, R. (1978): "The complex nature of infant speech perception", Proceedings of the conference on Child Phonology: Perception, Production and Deviation, G.H. Yeni-Komshian, J.F. Kavanagh and C.A. Ferguson (eds.) (In preparation).
- Fant, G. (1977): "Introductory remarks - Laryngeal mechanisms and features", in "The larynx and language", G. Fant and C. Scully (eds.), Phonetica 34, 252-255.
- Fant, G. and C. Scully (eds.) (1977): "The larynx and language", Phonetica 34, 247-325.
- Ferguson, C.A. (1978): "Learning to pronounce: The earliest stages of phonological development in the child", in Communicative and cognitive abilities - early behavioral assessment, F.D. Minifie and L.L. Lloyd (eds.), Baltimore, Md.: University Park Press (In press).
- Flanagan, J.L., K. Ishizaka, and K.L. Shipley (1975): "Synthesis of speech from a dynamic model of the vocal cords and vocal tract", Bell Syst. Tech. J. 54, 485-506.
- Folkins, J.W. and J.H. Abbs (1975): "Lip and jaw motor control during speech responses to resistive loading of the jaw", JSHR 18, 207-220.
- Fowler, C.A. (1977): Timing control in speech production, Bloomington, Ind.: Indiana University Linguistics Club.
- Fowler, C.A., P. Rubin, R.E. Remez, and M.T. Turvey (1978): "Implications for speech production of a general theory of action", in Language production, B. Butterworth (ed.), New York: Academic Press. (In press).
- Fromkin, V.A. (1973): Speech errors as linguistic evidence, The Hague: Mouton.
- Fry, D.B. (1964): "The functions of the syllable", Zs.f.Ph. 17, 215-237.
- Fujimura, O. (1977a): "Control of the larynx in speech", Phonetica 34, 280-288.
- Fujimura, O. (1977b): "Model studies of tongue gestures and the derivation of vocal tract area functions", in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.) 225-232, Tokyo: University of Tokyo Press.



- Fujimura, O. and Y. Kakita (1978): "Remarks on quantitative description of the lingual articulation", in Frontiers of speech communication research, B. Lindblom and S. Ohman (eds.), London: Academic Press.
- Galaburda, A.M., M. LeMay, T.L. Kemper and N. Geschwind (1978): "Right-left asymmetries in the brain", Science 199, 852-856.
- Gay, T. (1977): "Cinefluorographic and electromyographic studies of articulatory organization", in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.), 85-102, Tokyo: University of Tokyo Press.
- Geschwind, N. (1972): "Language and the Brain", Scientific American 226, 76-83.
- Ghazelli, S. (1977): Back consonants and backing coarticulation in Arabic, unpublished Ph.D. dissertation, University of Texas at Austin.
- Green, E. and D.H. Howes (1977): "The nature of conduction aphasia: A study of anatomic and clinical features and of underlying mechanisms", in Studies in Neurolinguistics, Vol. 3, H. and H.A. Whitaker (eds.), 123-156, New York: Academic Press.
- Greene, P.H. (1972): "Problems of organization of motor systems", in Progress in Theoretical Biology, Vol. 2, R. Rosen and F. Snell (eds.), 304-322, New York: Academic Press.
- Harshman, R., P. Ladefoged and L. Goldstein (1977): "Factor analysis of tongue shapes", JASA 62, 693-707.
- Henneman, E., G. Somjen and D.D. Carpenter (1965): "Functional significance of cell size in motoneurons", J. Neurophysiol. 28, 560-580.
- Hirano, M. (1977): "Structure and vibratory behavior of the vocal folds", in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.), 13-27, Tokyo: University of Tokyo Press.
- Hixon, T.J., J. Mead and M.D. Goldman (1976): "Dynamics of the chest wall during speech production: Function of the thorax, rib cage, diaphragm, and abdomen", JSHR 19, 297-356.
- Jakobson, R. (1941): Kindersprache, Aphasie, und allgemeine Lautgesetze, Uppsala: Almqvist and Wiksell. Reprinted as Child language, aphasia, and phonological universals (1968), The Hague: Mouton.
- Jusczyk, P.W., B.S. Rosner, J.E. Cutting, C.F. Foard and L.B. Smith (1977): "Categorical perception of nonspeech sounds by 2-month-old infants", Percept. Psychophys. 21, 50-54.
- Kent, R.D. and F.D. Minifie (1977): "Coarticulation in recent speech production models", JPh 5, 115-134.
- Kimura, D. (1976): "The neural basis of language qua gesture", in Studies in Neurolinguistics, Vol. 2, H. and H.A. Whitaker (eds.), 145-156, New York: Academic Press.
- Kimura, D. and Y. Archibald (1974): "Motor functions of the left hemisphere", Brain 97, 337-350.
- Kortlandt, A. (1973): Discussion of paper by G.W. Hewes, "Primate Communication and the gestural origin of language", Current Anthropology 14, 13-14.

- Ladefoged, P. (1967): Three areas of experimental phonetics, London: Oxford University Press.
- Ladefoged, P. (1975): A course in phonetics, New York: Harcourt Brace Jovanovich.
- Ladefoged, P. (1977): "The description of tongue shapes", in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.) 209-222, Tokyo: University of Tokyo Press.
- Ladefoged, P., J. DeClerk, M. Lindau, and G.A. Papçun (1972): "An auditory-motor theory of speech production", U.C.L.A. Working Papers in Phonetics 22, 48-75.
- Lashley, K.S. (1951): "The problem of serial order in behavior", in Cerebral mechanisms in behavior, L.A. Jeffress (ed.), The Hixon Symposium, New York: Wiley.
- Lassen, N.A., D.H. Ingvar, and E. Skinhøj (1978): "Brain function and blood flow", Scientific American 239, 62-71.
- Lieberman, P. (1972): Discussion in Proceedings of the Seventh International Congress of Phonetic Sciences, A. Rigault and R. Charbonneau (eds.), 272, The Hague: Mouton.
- Lieberman, P. (1975): On the origins of language, New York: Macmillan.
- Lindblom, B.E.F. (1963): "Spectrographic study of vowel reduction", JASA 35, 1773-1781.
- Lindblom, B.E.F. (1964): "Articulatory activity in vowels", STL-QPSR, 3.
- Lindblom, B., J. Lubker and T. Gay (1978): "Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation", JPh (in press).
- Lindblom, B.E.F. and J.E.F. Sundberg (1971a): "Acoustical consequences of lip, tongue, jaw and larynx movement", JASA 50, 1166-1179.
- Lindblom, B.E.F. and J. Sundberg (1971b): "Neurophysiological representation of speech sounds". Paper presented at the XVth World Congress of Logopedics and Phoniatrics, Buenos Aires, Argentina, August.
- Lisker, L., A.S. Abramson, F.S. Cooper, and M.H. Schvey (1969): "Transillumination of the glottis in running speech", JASA 45, 1544-1546.
- Lisker, L. and A.S. Abramson (1972): "Glottal modes in consonant distinctions", in Proceedings of the Seventh International Congress of Phonetic Sciences, A. Rigault and R. Charbonneau (eds.), 366-370, The Hague: Mouton.
- MacKay, D.G. (1970): "Spoonerisms: The structure of errors in the serial order of speech", Neuropsychologia 8, 323-350.
- MacNeilage, P.F. (1970): "Motor control of serial ordering of speech", Psych. Rev. 77, 182-196.
- MacNeilage, P.F. (1977): "Is the speaker-hearer a special hearer?", in Modes of perceiving and processing information, H.L. Pick and E. Saltzman (eds.), 53-66, New Jersey: L. Erlbaum Assoc.

- MacNeilage, P.F. (1978a): "The control of speech production", Proceedings of the conference on Child Phonology: Perception, Production and Deviation, G.H. Yeni-Komshian, J.F. Kavanagh and C.A. Ferguson (eds.). (In preparation).
- MacNeilage, P.F. (1978b): "Neural mechanisms in speech production", in Current issues in the phonetic sciences, H. and P.A. Hollien (eds.), Amsterdam: John Benjamins B.V. (In preparation).
- Mateer, C. and D. Kimura (1977): "Impairment of nonverbal oral movements in aphasia", Brain and Language 4, 262-276.
- Moll, K. (1977): "Physiological studies of speech production: Supralaryngeal", in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.), 404-408, Tokyo: University of Tokyo Press.
- Moll, K.L., G.N. Zimmerman and A. Smith (1977): "The study of speech production as a human neuromotor system", in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.), 107-128, Tokyo: University of Tokyo Press.
- Muller, E., J. Abbs, J. Kennedy and C. Larson (1977): "Significance of perioral biomechanics to lip movements during speech", paper presented to the American Speech and Hearing Association, Chicago.
- Nooteboom, S. (1970): "The target theory of speech production", IPO Annual Progress Report 5, 51-55.
- Ohala, J. (1974): "A mathematical model of speech aerodynamics", Proc. Speech Comm. Sem., Stockholm, 65-72, Uppsala: Almqvist and Wiksell.
- Oller, D.K. (1978): "The emergence of the sounds of speech in infancy", Proceedings of the conference on Child Phonology: Perception, Production and Deviation, G.H. Yeni-Komshian, J.F. Kavanagh and C.A. Ferguson (eds.). (In preparation).
- Oller, D.K., L.A. Wieman, W.J. Doyle, and C. Ross (1976): "Infant babbling and speech", Journal of Child Language 3, 1-11.
- Penfield, W. and L. Roberts (1959): Speech and brain mechanisms, Princeton, N.J.: Princeton University Press.
- Riordan, C.J. (1977): "Control of vocal-tract length in speech", JASA 62, 998-1002.
- Sawashima, M. and F.S. Cooper (eds.) (1977): Dynamic aspects of speech production, Tokyo: University of Tokyo Press.
- Searleman, A. (1977): "A review of right hemisphere linguistic capabilities", Psych. Bull. 84, 503-528.
- Shattuck-Hufnagel, S.R. and D.H. Klatt (1978): "Symmetry in the direction of substitution for segmental speech errors", JVLVB (in press).
- Stevens, K.N. (1977): Discussion in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.), 343-344, Tokyo: University of Tokyo Press.
- Stevens, K.N. and J.S. Perkell (1977): "Speech physiology and phonetic features", in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.), 323-341, Tokyo: University of Tokyo Press.
- Sussman, H.M., P.F. MacNeilage and R.J. Hanson (1973): "Labial and mandibular mechanics during the production of bilabial stop consonants", JSHR 16, 397-420.
- Titze, I.R. (1976): "On the mechanics of vocal-fold vibration", JASA 60, 1366-1380.
- Turvey, M. (1977): "Preliminaries to a theory of action with reference to vision", in Perceiving, acting and knowing: Toward an ecological psychology, R. Shaw and J. Bransford (eds.), Hillsdale, N.J.: Erlbaum.
- Turvey, M.T., R.E. Shaw, and W. Mace (1978): "Issues in the theory of action", in Attention and performance VII, J. Requin (ed.), Hillsdale, N.J.: Erlbaum. (In press).
- Ushijima, T. and H. Hirose (1974): "Electromyographic study of the velum during speech", JPh 2, 315-326.
- van den Berg, J. (1958): "Myoelastic-aerodynamic theory of voice production", JSHR 1, 227-244.
- Verbrugge, R.R., W. Strange, D.P. Shankweiler, and T.R. Edman (1976): "What information enables a listener to map a talker's vowel space?", JASA 60, 198-212.
- Warren, R.M. (1976): "Auditory perception and speech evolution", in Origins and evolution of language and speech, S.R. Harnad, H.D. Steklis and J. Lancaster (eds.), 708-717, New York: New York Academy of Sciences.
- Westbury, J. (1978): Aspects of the temporal control of voicing in consonant clusters in English, unpublished Ph.D. dissertation, University of Texas at Austin.



## ARTICULATORY PARAMETERS

Peter Ladefoged, Phonetics Lab., Linguistics Department, UCLA,  
Los Angeles, CA 90024, USA

The main report for this session gives an excellent summary of recent research on speech production. I would like to try to summarize this summary by listing and discussing the articulatory parameters that need to be controlled in a model of the speech production process. Obviously this could be done at various levels of generality. For example, one could choose to model the various muscular forces acting on the tongue, as suggested by Fujimura and Kakita (1978), or one could model the results of those forces as described by Harshman et al. (1977). Similarly one could specify the gross respiratory movements as Ohala (1975) has done, or more simply the variations in subglottal pressure that result from those movements. On another dimension of generality, one could try to describe just those articulatory parameters required for a particular language, or the larger set that would produce all possible linguistic differences, or even those that would go still further and allow one to distinguish all the personal characteristics of individual speakers.

I have chosen to specify speech production in terms of the minimal set of articulatory parameters given in Table 1. They will (hopefully) account for all linguistic differences both within and between languages, but may not distinguish between speakers. There is a lot of guess-work involved in setting up a list of this kind. Some of the parameters (eg 1, 2, 8, 9, 11, 16) can be defined fairly precisely, but others (eg 5, 6, 7, 14) are less firmly established.

The parameters listed may be thought of as corresponding to what is controlled rather than to movements of anatomical structures such as the jaw or the ribcage. This is a somewhat controversial point in that Lindblom and Sundberg (1971) have proposed that it is more appropriate to model tongue movements with respect to a moving mandible, rather than simply modeling the vocal tract shapes that result from these tongue movements. But it seems to me that if one is trying to state the parameters that are used in controlling articulatory actions, then Lindblom's own work (Lindblom et al., 1978) shows that speakers may rely on a great deal of

Table 1

A necessary and sufficient set  
of articulatory parameters.

- |                               |                           |
|-------------------------------|---------------------------|
| 1. Front raising              | 9. Lip width              |
| 2. Back raising               | 10. Lip protrusion        |
| 3. Tip raising                | 11. Velic opening         |
| 4. Tip advancing              | 12. Larynx lowering       |
| 5. Pharynx width              | 13. Glottal aperture      |
| 6. Tongue bunching            | 14. Phonation tension     |
| 7. Lateral tongue contraction | 15. Glottal length        |
| 8. Lip height                 | 16. Lung volume decrement |

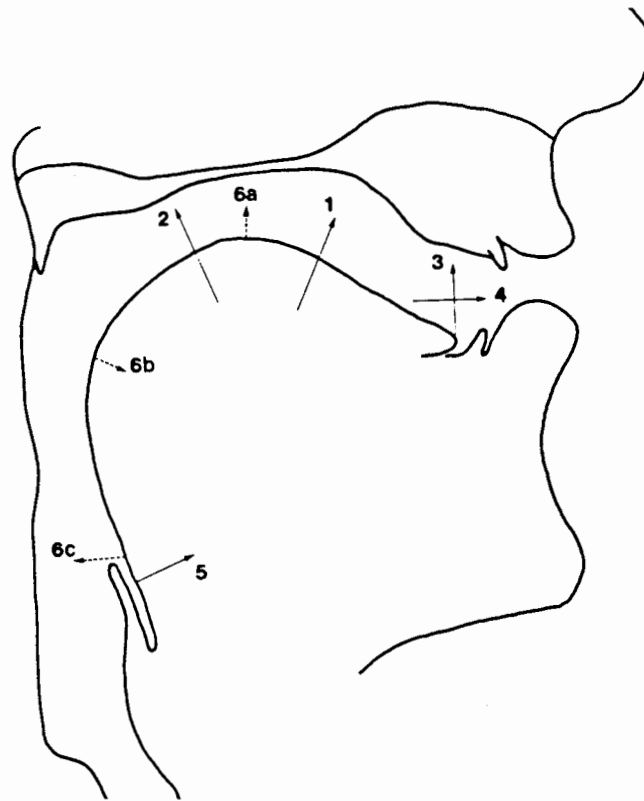


Figure 1

The movements of principal portions of the tongue associated with the first 6 parameters in Table 1.

compensation between movements of the jaw and those of the tongue. What they control are the vocal tract shapes, i.e. the relative magnitudes of the cross-sectional areas of the mouth and pharynx. The underlying parameters may therefore be as shown in Table 1.

The first six parameters are concerned with the position of the tongue relative to the roof of the mouth and the back wall of the pharynx. Most of these also involve movements of the soft palate and the pharynx, and it is only a convenient simplification to regard them as merely movements of the tongue. They are really parameters for the control of vocal tract shape.

For each of the first five parameters there is one portion of the tongue which makes the largest movement, and this portion may be used to name the parameter as a whole. These movements are shown in Figure 1.

It should be emphasized that each parameter specifies more than the movement of a single point. Thus the first parameter, front raising, specifies the degree of raising or lowering of the front of the tongue, and also the concomitant advancement or retraction of the root of the tongue. To say that a given sound has a certain degree of front raising means that the tongue as a whole may be said to be deviating from a neutral reference position to that degree. The arrow marked 1 in Figure 1 shows the potential movements of that part of the tongue that moves most with variations in front raising. Other points will move to a lesser degree.

The first two parameters, front raising and back raising (arrows 1 and 2), have been fully described in a series of recent publications (Harshman et al. 1977, Ladefoged et al. 1978, Ladefoged and Harshman 1979). These parameters enable us to give explicit formal descriptions of the movements of the tongue of an average speaker, such that we can characterize, fairly accurately, at least the non-rhotacized vowels of English.

It is obviously of interest to phoneticians to compare descriptions in terms of front raising and back raising with more traditional descriptions in terms of the highest point of the tongue, but unfortunately this cannot be done at the moment. The problem with these traditional descriptions is that no one has as yet shown how to interpret them unambiguously. Given the height and degree of backness of the highest point of the tongue (and given that all the other parameters such as pharynx width

have neutral values) it is not yet known how (or even if) the position of the tongue as a whole may be described.

The remaining parameters in Figure 1 have not been investigated as fully as the first two. It seems clear that there must be two degrees of freedom to movements of the tip of the tongue, as suggested by the arrows marked 3 and 4. There are many sounds which involve advancing or retracting the tip of the tongue while raising or lowering it in varying degrees. But we do not really know exactly what it is that is controlled, nor how these two parameters are related to one another. Furthermore, as Ohala (1974a) has pointed out, these movements may also affect the back of the tongue. It is impossible to do more than guess at a full mathematical specification of these parameters.

The fifth parameter, pharynx width, has been discussed extensively by Lindau (1979). For most languages, the position of the body of the tongue in vowels can probably be described very adequately in terms of the two parameters, front raising and back raising. But there are a number of languages such as Akan and Igbo, in which the width of the pharynx is independent of the height of the body of the tongue.

The three dotted lines in Figure 1 represent an estimate of the effect of the sixth parameter, tongue bunching. This estimate is based on an analysis of only five speakers of American English saying the vowel /ə/ as in "heard", and should be regarded as very tentative. Line 6a indicates a bunching up of the front of the tongue, 6b a concomitant increase in the opening of the vocal tract in the upper part of the pharynx, and 6c a considerable narrowing in the lower part of the pharynx. All these co-occur in tongue bunching in American English. But it should be noted that vowels of this kind are very unusual, and are likely to occur in less than 1% of the languages of the world (Maddieson, personal communication).

The final parameter associated with adjustments of tongue shape is lateral tongue contraction, which occurs in the production of laterals. Because the tongue is an incompressible mass, decreasing the lateral dimension must cause an increase in some other dimension. But we do not know how the narrowing movement is controlled. If speakers are aiming to control vocal tract shape, then decreases in tongue width may be complemented by

movements of the tongue within the mandible, absorbing potential increases in tongue height.

In addition to movements of the tongue (and the concomitant movements of the pharynx), there are a number of other parameters that affect the shape of the vocal tract. Foremost among these are movements of the lips. There are probably only three degrees of freedom involved: the distance between the upper and lower lip (lip height); the distance between the corners of the lips (lip width); and the degree of lip protrusion. In most languages the specifications of lip position in contrasting sounds do not require this number of degrees of freedom. But systematic phonetic differences between languages must also be taken into account. Thus French and German both have front rounded vowels, but there may be less lip protrusion in French.

The degree of velic opening is a well known parameter, and needs no further comment here. Similarly, it is well established that larynx raising and lowering is a controllable gesture that may occur in (among other sounds) different kinds of stop consonants.

There is more disagreement on the parameters required for characterizing glottal states. Despite the elaborate description of what is humanly possible that has been given by Catford (1977), it seems to me that languages use controllable differences in only three parameters: the distance between the arytenoid cartilages (glottal aperture), which is of course, the physiological parametric correlate of oppositions such as voiced-voiceless; the stiffness and mass of the parts of the vocal cords that may vibrate (glottal tension), which may be varied to produce different phonation types such as creaky voice; and the degree of stretching of the vocal cords (glottal length), which correlates most highly with the rate of vibration (the pitch).

The final parameter is lung volume decrement, the prime source of energy for nearly all speech sounds. This is highly correlated with the subglottal pressure, but should not be confused with it. It appears from the work of Ohala (1974) that speakers control the amount of work done by the respiratory system (the rate of decrease of lung volume), rather than the subglottal pressure. Thus they will produce a given amount of power for a given kind of word, irrespective of whether it contains a voiceless aspirate (which will

cause a fall in the subglottal pressure) or a glottal stop (which will cause an increase).

Most speech sounds have a unique specification in terms of these 16 parameters. MacNeilage's report may give a slightly wrong impression in this respect. It is not quite correct to say that "Ladefoged et al. (1972) showed that ... there is a considerable variation of tongue configurations adopted by different speakers producing the same vowel." We showed only that different speakers used different degrees of jaw opening to offset different degrees of movement of the tongue relative to the mandible. If by "tongue configurations" one means vocal tract shapes, then one can observe very few differences between speakers.

There are probably only two major ways in which variations in one parameter may lead to no change in the speech sound produced because they are offset by variations in another parameter. The first is the use of larynx lowering to offset decreases in lip rounding (Atal et al. 1977, Riordan 1977). The second is the use of increased respiratory power (lung volume decrement) to offset decreases in the stretching of the vocal cords (glottal length). There may also be variations among the three lip parameters that can be used to compensate for one another. But the data of Atal et al. (1977) on parameterized tongue shapes, and our own similar data, indicate that there are no cases in which a given sound can be produced with the same lip and larynx position, but with two different tongue shapes, as long as the tongue shape is characterized by only two parameters. There are well known cases involving additional parameters, such as American English rhotacized vowels that may be produced in two different ways (Uldall 1958). There may also be variations in pharynx width that can compensate at least in part for variations in front raising and back raising to produce similar tongue shapes in vowels. But apart from the case of rhotacized vowels, I doubt that there are two distinct tongue shapes that produce the same sound.

The 16 parameters listed are hypothesized to be a necessary and sufficient set for linguistic phonetic specifications. Some of them are far from fully defined, but they are all susceptible of precise numerical specification. They are potentially the things that are controlled in speech production. As MacNeilage indicates, we do not yet know whether speech production involves

specifying a sequence of targets or whether some form of action theory specification is preferable. The parametric approach outlined above is equally applicable in either case. Very tentatively, Table 1 is offered as a summary of what we use in speech production.

#### References

- Atal, B., J.J. Chang, M.V. Mathews, and J.W. Tukey (1978): "Inversion of articulatory to acoustic transformation by a computer sorting technique", JASA 63, 1535-1555.
- Catford, J.C. (1977): Fundamental Problems in Phonetics, Edinburgh: Edinburgh University Press.
- Fujimura, O. and Y. Kakita (1979): "Remarks on quantitative description of the lingual articulation", in Frontiers of Speech Communication Research, B. Lindblom and S. Ohman (eds.), London: Academic Press.
- Harshman, R., P. Ladefoged, and L. Goldstein (1977): "Factor analysis of tongue shapes", JASA 62, 693-707.
- Ladefoged, P., J. DeClerk, M. Lindau, and G.A. Papçun (1972): "An auditory-motor theory of speech production", UCLA Working Papers in Phonetics 22, 48-75.
- Ladefoged, P., R. Harshman, L. Goldstein, and D.L. Rice (1978): "Generation of vocal tract shapes from formant frequencies", JASA 64, 1027-1035.
- Ladefoged, P. and R. Harshman (1979): "Formant frequencies and movements of the tongue", Frontiers of Speech Communication Research, B. Lindblom and S. Ohman (eds.), London: Academic Press.
- Lindblom, B., J. Lubker, and T. Gay (1978): "Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation", JPh (in press).
- Lindblom, B.E.F. and J.E.F. Sundberg (1971): "Acoustical consequences of lip, tongue, jaw and larynx movement", JASA 50, 1166-1179.
- Lindau, M. (1979): "The feature Expanded", JPh (in press).
- Maddieson, I. (personal communication): "Phonological inventories of the languages of the world".
- Ohala, J. (1974): "A mathematical model of speech aerodynamics", Proc. Speech Comm. Sem., Stockholm, 65-72, Uppsala: Almqvist and Wiksell.
- Ohala, J. (1974a): "Phonetic explanation in phonology", Papers from the parasession on natural phonology, 251-271, Chicago: Chicago Linguistic Society.
- Riordan, C.J. (1977): "Control of vocal-tract length in speech", JASA 62, 998-1002.
- Uldall, E.T. (1958): "American 'molar' r and 'flapped' r", Revista do Laboratorio de Fonetica Experimental, Coimbra 4, 103-106.

## A SUPPLEMENTARY REPORT ON SPEECH PRODUCTION

Masayuki Sawashima, Research Institute of Logopedics and Phoniatrics, Faculty of Medicine, University of Tokyo, Tokyo, Japan.

In this paper, three different subtopics of speech production are discussed. They are: 1. Laryngeal control for voicing distinction. 2. Articulatory dynamics in normal and dysarthric cases. 3. Central mechanism of skilled movements.

1. Laryngeal Control for Voicing Distinctions

The basic features of laryngeal movements for the voiced vs. voiceless distinction in various languages have been examined by use of EMG and fiberoptic techniques. The glottal adduction-abduction dimension is directly observable with the fiberscope. The general picture is that the glottis is closed or nearly closed for voiced sounds while it is open for voiceless sounds, the extent of glottal opening varying with different phonemes and phonological environments. For Japanese voiceless stops the glottal opening for the same phoneme is greater in word-initial position than in word-medial position. For geminate stops, which occur only in word-medial position in Japanese, the duration of the glottal opening is consistently longer than for the corresponding non-geminate stops, whereas the degree of opening is often observed to be as small as in word-medial non-geminates. The findings for stops are also applicable to affricates. Voiceless fricatives show wide glottal aperture even in word-medial position. A large glottal aperture is associated with vowel devoicing in Japanese (Sawashima et al. 1976). In American English, a larger glottal opening associated with a greater degree of aspiration is observed for pre-stressed voiceless stops as compared to the corresponding post-stressed stops (Sawashima 1970). In French voiceless stops, a larger glottal opening is also observed for the pre-stressed position than for the post-stressed position (Benguerel et al. 1978). Observations on languages such as Korean (Kagaya 1974, Hirose et al. 1974), Hindi (Kagaya and Hirose, 1975) and Chinese (Iwata and Hirose, 1976), which have a phonemic distinction between aspirated and unaspirated stops, have revealed a large glottal opening for aspirated voiceless stops. The articulatory release takes place nearly at the point when the maximum glottal opening is reached. For unaspirated voiceless stops, on the other hand, the glottis is

nearly closed to the phonatory position at the articulatory release, although a small amount of glottal separation is observable during oral closure. A similar contrast has been observed between /p/ and /b/ in Danish in word-initial position (Fischer-Jørgensen and Hirose 1974a). For the voiced aspirated stops of Hindi (Kagaya and Hirose 1975), the glottis is closed during most of the oral closure until it begins to open at the end of the oral closure, the maximum opening being reached after the release.

It has been observed that the glottal stop gesture, instead of glottal abduction, is used for English voiceless stops in certain environments (Fujimura and Sawashima 1971). The characteristic appearance of this gesture is an adduction of the false vocal cords covering the closed glottis. In whispered phonation, there is a constriction of the supraglottal laryngeal structures characterized by the adduction of the false vocal cords and a reduction in the anteroposterior dimension of the laryngeal cavity, although the glottis is open as in voiceless sounds in normal speech (Weitzman et al. 1976). The adduction of the false vocal cords appears to contribute to prevent the glottal vibration by the transglottal air flow and also to facilitate the generation of turbulent noise.

Electromyographic study of the larynx (Hirose and Gay 1972, Hirose and Ushijima 1978, Hirose et al. 1978a) has revealed, in various languages, a clear reciprocal pattern of activity between the posterior cricoarytenoid (PCA) and interarytenoid (INT) muscles for the voiced vs. voiceless distinction, a decrease in PCA activity with an increase in INT activity for voiced sounds, and the reverse for voiceless sounds. It has also been revealed that PCA is important for active vocal fold abduction for those speech sounds which are produced with an open glottis (Hirose 1976, Hirose and Ushijima 1978, Hirose et al. 1978a). A detailed observation of the laryngeal movements in correspondence with the EMG patterns for various types of Japanese voiceless sounds and sound sequences (Sawashima et al. 1978) has revealed that there is some subject-to-subject difference in the mode of laryngeal control using the PCA and INT muscles. In one subject, the time course and the extent of the glottal aperture are mainly represented by PCA activity with an associated decrease of INT activity. The time curve of the glottal width in this case can be interpreted as a kind of mechanically smoothed pattern of PCA activity. In the other subject,

however, the activity of the INT appears to actively contribute, in combination with the PCA, to the control of the glottal condition.

The data mentioned above present fairly clear physiological evidence for the laryngeal control of the glottal abduction and adduction. Another problem is whether or not we see physiological evidence for the stiff-slack dimension of the vocal folds, which was proposed by Halle and Stevens (1971) as another laryngeal feature contributing to the voiced-voiceless distinction in addition to abduction-adduction of the glottis. According to their proposal, stiffening of the vocal folds takes place for voiceless consonants and slackening for voiced consonants. When considering the physiological mechanism of control of vocal fold stiffness, we should refer to the "cover and body" structure of the vocal folds proposed by Hirano (1974). According to Hirano, the vocal folds consist of two different layers which are connected loosely with each other. The outer layer, which is called the "cover", is the mucosa covering the free edge of the vocal folds. The inner layer, which is called the "body", contains the vocalic muscle. The longitudinal pull of the vocal folds by the contraction of the cricothyroid (CT) muscle or some other external force results in an increase in the stiffness of both the cover and the body. Contraction of the vocalis (VOC) muscle also causes stiffening of the body, but it may shorten the vocal folds and result in slackening of the cover which would facilitate the vocal fold vibration (Fujimura 1977). Thus increase in the activity of CT can be physiological evidence for stiffening of the vocal folds, although other possible mechanisms are still to be explored. An extensive EMG study of the role of the larynx for the voicing distinction in Japanese consonants (Hirose and Ushijima 1978), has revealed that there is a temporary decrease in CT activity for both voiced and voiceless consonants. The degree of suppression is greater for word-initial voiced consonants than for the voiceless counterparts and least for word-medial consonants with no difference according to the voicing condition. There was also a temporary suppression of the VOC activity, the extent of the suppression being independent of the voicing condition but greater for the word-initial consonants than for the word-medial sounds. The results reveal that in Japanese consonants there is no physiological evidence for

the stiff-slack dimension in the laryngeal control of the voicing distinction. The greater suppression of the muscle activities is considered to be related to the  $F_0$  fall and the presence of the word boundary. In Hindi, Dixit (1975) reported a high CT activity for voiceless stops, but the results of Kagaya and Hirose (1975) for the same language failed to confirm that. A higher level of CT activity and a lower level of VOC activity for the voiceless stops than for the voiced stops is reported in a study of Swedish short and long consonants (Hirose 1977). Hirose et al. (1974) reported a sharp increase in VOC activity immediately before the articulatory release of a Korean forced stop. This particular VOC activity was interpreted as a physiological correlate of laryngealization as observed in the Danish stød (Hirose et al. 1974, Fischer-Jørgensen and Hirose 1974b, Fischer-Jørgensen 1977, Hirose et al. 1978a). Another interpretation proposed by Fujimura (1977, 1978) is that the vocalis muscle functions as a relatively fast-response voicing trigger mechanism for facilitating the vibration of the vocal folds which are otherwise under unfavorable conditions because of their tenseness.

In summary, physiological correlates of the tenseness feature appear to be manifested in some of the experimental results, but they are not as clear and consistent as those of the adduction-abduction dimension of the laryngeal features. It is reasonable to assume, however, that the laryngeal adjustments for the voicing distinction are not limited to simple adduction-abduction of the vocal folds. Further study is needed to explore the physiological correlates of some other features including the problem of tenseness of the vocal folds.

## 2. Articulatory Dynamics in Normal and Dysarthric Cases

In studying dynamic aspects of the articulatory movements, various basic characteristics such as the velocity of movement in different parts of the speech organs should be taken into account. Analysis of the articulatory movements in the repetitive production of a monosyllable is considered to present valuable information in this respect. According to Hudgins and Stetson (1937), the maximum rate of syllable repetition (mean value per second) is: 6.7 for the lip in /pu/, 8.2 for the tip of the tongue in /tat/, 7.1 for the back of the tongue in /ka/, and 6.7 for the velum in /tun/. Recordings of the actual movements and EMG of the relevant

muscles provide data not only on the velocity but also on other aspects such as accuracy of the movement, regularity of the rhythmic pattern and muscle coordination. The development of the X-ray microbeam system (Kiritani et al. 1975) enabled us to make detailed analyses of the articulatory movements in normal subjects and also in patients with various neuromotor disorders (Hirose et al. 1978b, Hirose et al. 1978c).

The maximum velocity (mm/sec) in the syllable repetition for a normal subject is: 190-250 for the lip in /pa/, 220 for the tip of the tongue in /t/ of /pataka/, 200-220 for the back of the tongue in /ka/, and 105 for the velum in /teN/. It is noted that the velocity of the velar movement is definitely slower than the others. In the normal subject, the repetition of the movement is carried out quite regularly in terms of the amplitude, velocity, interval and also direction of the movement. In the normal subject it is also noted that attempts to make the syllable repetition at a slower rate do not result in a decrease in velocity but in an increase in the closure period of the given consonant as compared to a faster rate of repetition. EMG of the pertinent articulatory muscles shows a quite regular rhythmic pattern of activation-suppression corresponding to the movement with a clear reciprocal activity pattern between the antagonistic muscle pairs. The syllable repetition by patients with amyotrophic lateral sclerosis (ALS) is characterized by a slow rate of repetition and a decrease in both the velocity and amplitude of the movement, while the regularity of the movement is maintained. Patients with cerebellar ataxia are characterized by an irregular fluctuation of the interval, velocity and amplitude of the movement in syllable repetition. Electromyographic patterns also reveal irregularity in both the extent and timing of muscle activation. There is a plateau during the period of suppression indicating a disturbance of initiation of muscle activity in repetitive movements. Reciprocity between the antagonistic muscles is somehow preserved. The abnormal characteristic pattern of patients with Parkinsonism is: a small range of amplitude with a repetition rate comparable to normal. In addition, the amplitude gradually decreases throughout the repetition series until the movement finally stops. A gradual decrease in the velocity is also characteristic. Electromyographic records reveal a regular pattern of activation-suppression in each muscle, but the

temporal reciprocity between the antagonistic muscles is not maintained and the two muscles are rather synchronously activated.

The dynamic characteristics presented here appear to well reflect the underlying motor pattern of voluntary movements in normal and various types of pathologic conditions. Thus the analysis of the syllable repetition is a promising approach for a differential diagnosis of various types of dysarthrias as well as for the study of dynamic aspects of speech production.

### 3. Central Mechanism of Skilled Movements

The central mechanism of dynamic adaptive motor control in speech production has been discussed by many researchers. One thing to be kept in mind here is the fact that the articulatory movements, although speech specific, are a kind of learned skilled voluntary movements. In this sense, it would be useful to refer to the central mechanism of other skilled movements which was suggested by Allen and Tsukahara (1974). They discussed the participation of the cerebellum in the planning and carrying out of a voluntary movement of the limbs in their extensive review on the cerebro-cerebellar communication systems. According to them, the most reasonable possibility for the lateral cerebellum is that it participates in the programming or long-range planning of the movement. The intermediate cerebellum works as a feedback system to the motor cortex in the execution of the movement. They state:

"Once the movement has been planned within the association cortex, with the help of the cerebellar hemisphere and basal ganglia, the motor cortex issues the command for movement. At this point the pars intermedia (of the cerebellum) makes an important contribution by updating the movement based on the sensory description of the limb position and velocity on which the intended movement is to be superimposed. This is a kind of short-range planning as opposed to the long-range planning of the association cortex and lateral cerebellum.... In learning a movement, we first execute the movement very slowly because it cannot be adequately preprogrammed. Instead, it is performed largely by cerebral intervention as well as by constant updating of the intermediate cerebellum. With practice, a greater amount of the movement can be preprogrammed and the movement can be executed more rapidly (without reference to peripheral sensory input." Thus, for learned movements the cerebellum provides an internal substitute for the external



world. "This cerebellar operation we consider to take place in the lateral zone." Although they discuss only the control of limb actions, the main points may also be applied to the speech actions. In studying speech dynamics, we should refer to a more general physiological basis of development and organization of the skilled movements on one hand, and explore various speech specific problems on the other hand.

#### References

- Allen, G.L. and N. Tsukahara (1974): "Cerebrocerebellar communication system", Physiol. Rev. 54, 957-1006.
- Benguereel, A-P., H. Hirose, M. Sawashima and T. Ushijima (1978): "Laryngeal control in French stop production: a fiberoptic, acoustic and electromyographic study", Folia Ph. 30, 175-198.
- Dixit, R.P. (1975): "Neuromuscular aspects of laryngeal control: with special reference to Hindi", Ph.D. Dissertation, the Univ. Texas at Austin.
- Fischer-Jørgensen, E. (1977): Discussion in Dynamic Aspects of Speech Production, M. Sawashima and F.S. Cooper (eds.) 70, Univ. Tokyo Press.
- Fischer-Jørgensen, E. and H. Hirose (1974a): "A preliminary electromyographic study of labial and laryngeal muscles in Danish stop consonant production", Haskins SR-39/40, 231-254.
- Fischer-Jørgensen, E. and H. Hirose (1974b): "A note on laryngeal activity in the Danish "stød"", Haskins SR-39/40, 255-259.
- Fujimura, O. (1977): "Control of the larynx in speech", Phonetica 34, 280-288.
- Fujimura, O. (1978): "Physiological functions of the larynx in phonetic control", Paper at the IPS-77, Miami.
- Fujimura, O. and M. Sawashima (1971): "Consonant sequences and laryngeal control", Ann. Bull. RILP 5, 1-6.
- Halle, M. and K.N. Stevens (1971): "A note on laryngeal features", Q. Prog. Rep. Res. Lab. Electron., MIT 101, 198-213.
- Hirano, M. (1974): "Morphological structure of the vocal cords as a vibrator and its vibrations", Folia Ph. 26, 89-94.
- Hirose, H. (1976): "Posterior cricoarytenoid as a speech muscle", Ann. Otol. 85, 334-343.
- Hirose, H. (1977): "Electromyography of the larynx and other speech organs", in Dynamic Aspects of Speech Production, M. Sawashima and F.S. Cooper (eds.), 49-65, Univ. Tokyo Press.
- Hirose, H. and T. Gay (1972): "The activity of the intrinsic laryngeal muscles in voicing control: an electromyographic study", Phonetica 25, 140-164.
- Hirose, H., C.Y. Lee and T. Ushijima (1974): "Laryngeal control in Korean stop production", JPh. 2, 145-152.
- Hirose, H. and T. Ushijima (1978): "Laryngeal control for voicing distinction in Japanese consonant production", Phonetica 35, 1-10.

- Hirose, H., H. Yoshioka and S. Niimi (1978a): "A cross language study of laryngeal adjustment in consonant production", Ann. Bull. RILP 12, 61-71.
- Hirose, H., S. Kiritani, T. Ushijima and M. Sawashima (1978b): "Analysis of abnormal articulatory dynamics in two dysarthric patients", JSHD 43, 96-105.
- Hirose, H., S. Kiritani, T. Ushijima, H. Yoshioka and M. Sawashima (1978c): "Patterns of dysarthric movements in patients with Parkinsonism", Ann. Bull. RILP 12, 73-85.
- Hudgins, C.V. and R.H. Stetson (1937) cited from Suprasegmentals, I. Lehiste, MIT Press, 1970.
- Iwata, R. and H. Hirose (1976): "Fiberoptic acoustic studies of Mandarin stops and affricates", Ann. Bull. RILP 10, 47-60.
- Kagaya, R. (1974): "A fiberoptic and acoustic study of the Korean stops, affricates and fricatives", JPh. 2, 161-180.
- Kagaya, R. and H. Hirose (1975): "Fiberoptic, electromyographic and acoustic analysis of Hindi stop consonants", Ann. Bull. RILP 9, 27-46.
- Kiritani, S., K. Itoh and O. Fujimura (1975): "Tongue-pellet tracking by a computer-controlled X-ray microbeam system", JASA 57, 1516-1520.
- Sawashima, M. (1970): "Glottal adjustments for English obstruents", Haskins SR-21/22, 180-200.
- Sawashima, M., H. Hirose and S. Niimi (1976): "Glottal conditions in articulation of Japanese voiceless consonants", XVI Int. Congr. Logopedics and Phoniatics, Interlaken 1974, 409-414, Basel: Karger.
- Sawashima, M., H. Hirose and H. Yoshioka (1978): "Abductor (PCA) and adductor (INT) muscles of the larynx in voiceless sound production", Ann. Bull. RILP 12, 53-60.
- Weitzman, R.S., M. Sawashima, H. Hirose and T. Ushijima (1976): "Devoiced and whispered vowels in Japanese", Ann. Bull. RILP 10, 61-80.

## S p e e c h   P e r c e p t i o n

Report:	MICHAEL STUDDERT-KENNEDY	59
Co-Report:	LUDMILLA A. CHISTOVICH	
	Auditory processing of speech	83
Co-Report:	HIROYA FUJISAKI	
	Some remarks on recent issues in speech perception research	93

The reports will also be published in *Language and Speech*, 1980.

## SPEECH PERCEPTION

Michael Studdert-Kennedy, Queens College of the City University of New York and Haskins Laboratories, New Haven, Connecticut

The past few years of research in speech perception have been very active. The old questions are still there -- What are the units? How do we segment? Where are the invariants? -- but some old answers have turned out to be wrong and some new ones are beginning to emerge. The intricate articulatory and acoustic structure of the syllable is still at the center of the maze, but other sources of information for the listener -- prosody, syntax, semantics -- have begun to receive experimental attention: Studies of fluent speech are taking their place beside the established methods of syllable analysis and synthesis. Theory has dropped into the background (or perhaps the back room) and no one seems very eager to argue the merits of analysis-by-synthesis or the "motor theory" any more. Certainly, theory continues to guide research, but a refreshing atheoretical breeze has been blowing in from artificial speech understanding research (Klatt, 1977, in press, a) and from developmental psychology (Aslin and Pisoni, in press). In the latter regard, I shall not have much to say directly about infant speech perception, but much of what I have to say will bear on it. The infant is a listener, a very attentive one, because by learning to listen it learns to speak. In my opinion, only by carefully tracking the infant through its first two years of life shall we come to understand adult speech perception and, in particular, how speaking and listening establish their links at the basis of the language system. This said, let us begin, as infants do, with prosody.

Prosody

Prosody refers to the melody, rhythm, rate, amplitude, quality and temporal organization of speech. There has been an upsurge of interest in these factors in recent years, partly because they seem to hold a key to improved speech synthesis, partly because prosodic contributions to speech perception have been unjustly neglected (Cohen and Nooteboom, 1975; Nooteboom, Brokx and de Rooij, 1976). To say that prosody "contributes" to speech perception may seem to imply that speech perception is confined to segmental processes, of which prosody is a mere subsidiary, conveying no distinctive information of its own. This, of course, is false. Prosody carries much of that important indexical information (Abercrombie, 1967) without which, if it is dark, you don't know who is talking to you or whether he means what he says. However, it is with the adjunct functions -- contributions to segmental perception -- that I am concerned here.

One prosodic function is to maintain a coherent auditory signal. Darwin (1975) asked listeners to shadow a sentence on one ear, while a competing sentence was led into the other. At some arbitrary point, prosodic contours were suddenly switched across ears, while syntactic and semantic sequences were maintained. Prosodic continuity then often overrode syntax, semantics and ear of entry, leading to the intrusion of words from the supposedly unattended ear. Evidently, listeners were tracking the prosodic contour, a process that Nootboom et al. (1976) suggest may be necessary to maintain "perceptual integrity".

What physical dimensions of the signal sustain this integrity? Rate is probably not important, because quite sharp rate variations are regularly used to convey syntactic information (e.g., Klatt, 1976). Of course, rate can affect segmental classification (Ainsworth, 1972), but listeners adjust rapidly, within less than a second (Fujisaki, Nakamura and Imoto, 1975; Summerfield, 1975; Nootboom, et al., 1976). Amplitude changes, within limits, are also probably of little importance (Darwin and Bethell-Fox, 1977). In fact, the principal determinants of prosodic continuity seem to be fundamental frequency ( $F_0$ ) and spectrum: Nootboom et al. (1976) showed that, when pitches, alternating over a 2-6 Hz range, are imposed on a sequence of three vowels, repeated at intervals of less than 150 msec., the vowels split into two streams, as though from two speakers. The effect is reduced, if the vowels are granted a degree of spectral continuity by being placed into consonantal context. This work, taken with similar studies by Dorman, Cutting and Raphael (1975) and by Darwin and Bethell-Fox (1977), leads to the conclusion that continuity of both formant structure and  $F_0$  underlies the perceptual integrity of running speech.

A second prosodic function is to facilitate phrasal grouping. Here the main variables seem to be  $F_0$  and segment duration. Several studies have documented syntactic control of timing and segment duration in production (e.g. Cooper, 1976; Klatt, 1976). Klatt and Cooper (1975) show, further, that listeners expect segment duration to vary with the syntactic position of a word in a sentence. For example, they judge lengthened syllables to be more natural at the end of a clause than at the beginning or middle. Similarly, Nootboom et al. (1976) report that listeners judge a vowel of a particular length to be shorter if it occurs at the end of a word than if it occurs at the beginning. Presumably, such observations reflect listeners' habitual use of phrase-final lengthening as an aid to parsing.

The role of  $F_0$  has been more extensively studied. For example, Collier and 't Hart (1975) constructed synthetic utterances consisting of 13 or 15

200 msec steady-state, vowel-like "syllables", separated by 50 msec silent intervals. They imposed ten theoretically derived  $F_0$  contours ('t Hart and Cohen, 1973) on these syllables, deploying characteristic "continuation rises" and "non-final falls" to delimit the ends and the beginnings, respectively, of possible syntactic constituents. Finally, following Svensson (1974) and Kozhevnikov and Chistovich (1965), they asked listeners to write down syntactically acceptable sentences to match each contour in number of syllables, location of stresses and overall intonation. Of the resulting sentences, 72% matched the predicted syntactic structures. Since two hypotheses were under test here -- both the correctness of the theoretically derived contours and the listeners' capacity to infer syntactic structure from intonation -- this is a remarkably high score.

Finally, a third perceptual function of prosody has aroused a great deal of interest in recent years. This is the function -- nobody knows what it is -- supposedly fulfilled by rhythm. Martin (1972) wrote a persuasive paper in which he argued that speaking involves more than a simple concatenation of motor elements: like other motor behaviors speech is compelled, by natural constraints on the relative timing of components, to be rhythmic. Moreover, some components (syllables) are "accented", and these are predictable: accent level (or stress) covaries with timing and the main accents are equidistant (i.e., isochronous). Finally, since "...speaking and listening are dynamically coupled rhythmic activities..." (p. 489), listeners can predict the main stresses and can use that fact to "cycle" their attention, saving it, as it were, for the more important words.

There is, in fact, evidence from phoneme-monitoring experiments that reaction time (RT) is shorter to initial phonemes in stressed words than in unstressed (Shields, McHugh and Martin, 1974). This is apparently not due to the greater energy of the stressed words, since, if the words are presented in isolation, no RT difference appears (Shields, et al., 1974). Moreover, Cutler (1976) has found that the RT difference holds, even if stress, or the lack of it, is merely "predicted" by prior prosodic contour and if the actual target is acoustically identical in both conditions. Cutler and Foss (1977), demonstrate, further, that the RT advantage is not due to syntactic form class, since it is found for stressed function words as well as for stressed content words. They conclude that the reduced reaction time may reflect heightened attention to the semantic focus of a sentence, and they cite unpublished evidence from Allen and O'Shaughnessy that "...reliable correlates of semantic focus are to be found in the fundamental frequency contour (p. 10)."

By this last point Cutler and Foss seem to be cutting themselves free

from Martin's (1972) claim for isochrony, whether wisely or not remains to be seen. Lehiste (1977) has recently reopened the isochrony issue in a paper summarizing much of her research on the topic. She concludes that although isochrony is "primarily a perceptual phenomenon" (p. 253), it does have some basis in production and is therefore available for communicative use. Lehiste shows that English interstress intervals are often lengthened to signal a syntactic boundary.

Isochrony has also come under experimental scrutiny. Morton, Marcus and Frankish (1976), recording a list of spoken digits for experimental use, discovered that acoustically (onset to onset) isochronous sequences sounded anisochronous. Moreover, listeners, asked to adjust a sequence to perceptual isochrony, made it acoustically anisochronous. Morton, et al. (1976) coined the term "perceptual centers" ("P-centers") to refer to those points in a sequence of words that are equidistant when the words sound isochronous. But they were unable to locate the points or specify their acoustic correlates. Surprisingly, the P-center does not correspond to any obvious acoustic marker, such as sound onset, vowel onset or syllable peak. However, Fowler (Ms. submitted for publication) has recently discovered that "...when asked to produce isochronous sequences, talkers generate precisely the acoustic anisochronies that listeners require in order to hear a sequence as isochronous." The acoustic anisochronies apparently arise because the articulatory onsets of words beginning with sounds from different manner classes have acoustic consequences at different relative points in time. From a review of her own and related studies (e.g., Allen, 1972; Lindblom and Rapp, 1973), Fowler concludes that "...listeners judge isochrony based on acoustic information about articulatory timing rather than on some articulation-free acoustic basis." Finally, although this work seems to be a thread that might unravel isochrony, Fowler is cautious in her claims. Most of the relevant experimental studies have used monosyllables and artificially repetitive utterances. What inroads this approach can make into the apparent isochrony of phonetically heterogeneous running speech remains to be seen.

#### Segmentation and Invariance

We turn now from the broad questions of prosody to the narrower puzzle of the syllable on which the prosody is carried. In what follows, I assume (together with most other investigators) that our task is to understand the process by which phonemes or features are extracted from the signal. Let us begin with a question raised by Myers, Zhukova, Chistovich and Mushnikov (1975): Is segmentation an auditory process, preceding phonetic classification, or an automatic consequence of classification itself? Several studies

from the Pavlov Institute in Leningrad speak to the question. Chistovich, Fyodorova, Lissenko and Zhukova (1975) showed that a sudden amplitude drop, roughly in the middle of a 460 msec steady-state vowel, caused listeners to hear either two vowels or a VCV sequence, depending on the magnitude and rate of the amplitude decrease. Subsequently, Myers, et al. (1975) used an ingenious dichotic technique to suggest that such amplitude decreases are registered by the peripheral auditory system; they inferred that, since classification is presumably central, segmentation must precede classification. Finally, Zhukova, Zhukova and Chistovich (1974) reported on the use of a similar technique to study the effects of spectral variation at segment boundaries. The investigators presented a time-varying value of F2 (roughly 2200 to 800 Hz over 200 msec), to one ear, steady-state values of F1 and F3 to the other. The latter were interrupted by a 12-15 msec pause, of which the position could be set by the subject so as to vary the fused percept from hard to soft [r], that is, from [iru] to [ir'u]. Subjects reliably set the pause so that its endpoint coincided with an F2 value of roughly 1600 Hz. Since this value is close to that of the hard-soft boundary previously determined for the steady-state isolated consonants [s] and [s'], the authors infer that listeners were also judging the soft consonant [r'] by its F2 value at onset. They conclude that "the auditory system interprets the acoustic flow as a sequence of time segments between instants of variation" (p. 237), and that it derives consonantal information by sampling formant frequencies at these instants.

However, this conclusion does not seem to be forced by the data. On the one hand, the presumed peripheral segment boundary, determined by a sharp amplitude drop, seems to have something in common with the boundary proposed by certain automatic recognition procedures for isolating syllables rather than phonemes (e.g. Mermelstein, 1975). On the other hand, an invariant formant onset is not incompatible with the use of formant movement into the following vowel as a consonantal cue (see Dorman, Studdert-Kennedy and Raphael, 1977). My inclination therefore is to suppose that the preliminary auditory segmentation (if any) is syllabic rather than phonemic, and that within-syllable segmentation may often be synonymous with classification. I will return to this point below.

The view of the perceptual process, proposed by the Russian group, as a succession of brief time slices (rather than as the active continuous tracking suggested by studies of prosody), is close to that currently being explored by K.N. Stevens. In a succession of publications over recent years, Stevens (e.g. 1975) has elaborated on the "quantal nature of speech." He points out that, although the vocal apparatus is capable of producing a wide variety of sounds,

relatively few are actually used in the languages of the world. He attributes this restriction to a nonlinear relation between articulatory and acoustic parameters: some articulatory configurations are acoustically stable, in the sense that small changes in articulation have little acoustic effect, others are unstable in the sense that equally small changes have a substantial effect. The universal set of phonetic features is drawn from those articulatory configurations that generate acoustically stable, invariant "properties." The properties, it should be stressed, are higher order spectral configurations, rather than isolated cues such as F2 onset frequency. To define these configurations, Stevens has largely relied on computations from a vocal tract model. Finally, to assure quantal (or categorical) perception of the invariant properties and to afford the human infant a mechanism for netting them in the speech stream, Stevens postulates a matching set of innate "property detectors."

Empirical tests of the quantal theory have been few. But a recent study of English stops (Blumstein and Stevens, in press) is a good illustration of the approach, since it deals with a notoriously context-dependent set of sounds. The goal was to demonstrate the presence of invariant properties in the acoustic signal, sufficient for recognition by fixed templates. The first step was to record two male speakers reading random lists of the voiced stops [b d g], followed by each of five vowels [i e a o u]. Short-time spectra were then determined, integrated over a 26 msec window at onset. The spectra were used to construct, by trial and error, a template fitted to each place of articulation, such that it either correctly accepted or correctly rejected the majority of utterances. Descriptions of the templates ("diffuse-rising" for alveolar, "diffuse-falling" for labial and "compact" for velar) recall the terminology of distinctive feature theory.

In the second part of the study, a corpus of utterances was collected for classification by the templates. Six subjects (4 males, 2 females) recorded five repetitions each of the voiced and voiceless stop consonants [b d g p t k], followed by each of the vowels [a e i o u], or preceded by each of the vowels [i e a a u]. The resulting 1800 utterances were then analyzed spectrally in the same way as the original utterances, and compared with the templates. The results were: at least 80% (and often higher) correct rejection and correct acceptance for initial stops, a slightly lower performance for released final stops, although for some unreleased final stops scores dropped as low as 40%. Analysis of variance revealed significant differences in template matching performance as a function of vowel context, but performance was significantly above chance in every case. Quite similar

results have been reported by Searle, Jacobson and Rayment (1978) using a very much longer time slice (100-200 msec) and deriving their invariant patterns from a running sequence of spectra.

Where then does this leave us? 80% or better is a good score -- although, as A.M. Liberman has suggested to me, we might do almost as well with the binary recipe proposed by Cooper, Delattre, Liberman, Borst and Gerstman in 1952: high burst, falling F2 transition for alveolar; low burst, falling F2 transition for velar; low burst, rising F2 transition for labial.

The question, of course, is: Is this really the way that humans do it? Dorman et al., (1977), modeling their study on the work of Fischer-Jørgensen (1972), edited release bursts and/or formant transitions out of English voiced stop consonants ([b d g]), spoken before nine different vowels. Acoustic analysis of the bursts for a given place of articulation showed them to be largely invariant (cf. Zue, 1976). However, the bursts were not invariant in their effect: for the most part, listeners only perceived the bursts correctly, if their main spectral weight lay close to the main formant of the following vowel, as Stevens himself has suggested (1975, pp. 312-313). Kuhn (1975) has shown that the main vowel formant varies with the length of the cavity in front of the point of maximum tongue constriction. Since front cavity length is a function of place of articulation, an estimate of front cavity resonance is tantamount to an estimate of place of articulation. Thus, proximity on the frequency scale may facilitate perceptual integration of the burst with the vowel, enabling the listener to track the changing cavity shape characteristic of a particular place of articulation followed by a particular vowel.

Stevens (see especially, 1975) does not deny that contextually variable cues -- such as formant transitions, voice onset time, vowel formant structure -- can be used by the human listener. However, he regards them as "secondary," learned cues, acquired by repeated association with the "primary" invariant properties, and used only as safety devices when invariant cues fail. Given the many knotty questions concerning the possible mechanisms for extracting and interpreting these "secondary" context-dependent cues, one may wonder how an organism whose primary endowment is a set of passive templates learns to use them at all.

The question becomes even more pressing when one considers that there is no independent evidence for the existence of the hypothesized templates or property detectors. To understand this we must briefly review recent findings in the study of categorical perception.

Categorical perception

As is well known, early work with speech synthesizers showed that a useful procedure for defining the acoustic properties of a phoneme was to construct tokens of opponent categories, distinguished on a single phonological feature, by varying a single acoustic parameter along a continuum (e.g., [ba] to [da], [da] to [ta], etc.). If listeners were asked to identify these tokens, they tended to identify any particular stimulus in the same way every time they heard it: there were few ambiguous tokens. Moreover, if they were asked to discriminate between neighboring tokens, they tended to do very badly, if they assigned the two tokens to the same class, very well if they assigned them to different classes -- even though the acoustic distance between tokens was identical in the two cases. This phenomenon was dubbed "categorical perception" (Liberman, Harris, Hoffman and Griffith, 1957). Although there were usually no grounds for supposing that the acoustic variations along synthetic continua mimicked the intrinsic allophonic variations of natural speech, categorical perception in the laboratory was taken to reflect a necessary aspect of normal speech perception, namely, the rapid transfer of speech sounds into a phonetic or phonological code. The phenomenon was also believed by some people, including myself, to be peculiar to speech (Studdert-Kennedy, Liberman, Cooper and Harris, 1970).

However, we now know that categorical perception, as observed in the laboratory, is neither peculiar nor necessary to speech. Demonstrations that it is not peculiar we owe to Cutting and Rosner (1974) (rise-time at the onset of sawtooth waves, analogous to a fricative-affricate series); to Miller, Wier, Pastore, Kelly and Dooling (1976) (noise-buzz sequences analogous to the aspiration-voice sequences of a voice onset time (VOT) series); to Pisoni (1977) (relative onset time of two tones); and to Pastore, Ahroon, Baffuto, Friedman, Puleo and Fink (1977). These last investigators extended their work into vision, demonstrating categorical perception of critical flicker, with a sharp boundary at the flicker-fusion threshold. They also induced clearly categorical perception of a sine-wave intensity series by providing listeners with a constant-reference tone, or "pedestal," at the center of the series. Pastore et al. (1977) conclude that a continuum may be categorically divided either by a sensory threshold (as in flicker-fusion) or by an internal reference (as in the intensity series). Presumably, the portion of the signal with the earlier onset serves as a reference in a VOT series, while in a place of articulation series, cued by direction and extent of formant transitions, a reference is provided by the fixed vowel. If this last point is correct, we perceive a place series categorically precisely because the consonants are judged relationally rather than absolutely -- an interpretation not compatible

with the notion of invariant property detectors.

Just how an internal reference suppresses discrimination within categories is not clear, but the results of Carney, Widin and Viemeister (1977) suggest that it may simply serve to divert the listener's attention from other stimuli in the series. To Carney et al. (1977) (see also Pisoni and Lazarus, 1974; Samuel, 1977) we owe the demonstration that a VOT continuum need not be perceived categorically. Each of their subjects displayed good within-category discrimination after moderate training on a bilabial VOT continuum. Indeed, discrimination was so good that subjects were able to shift category boundaries on request and assign consistent labels to arbitrary subsets of the stimuli. The outcome suggests that "...utilization of acoustic differences between speech stimuli may be determined primarily by attentional factors, ...distinct from the perceptual capacities of the organism" (Carney, et al., p. 969).

This is precisely what is suggested by the numerous instances in which speakers of different languages perceive an acoustic continuum in different ways. (For a thorough review, see Strange & Jenkins, 1977). For example, while American English speakers perceive an [r] to [l] continuum categorically, Japanese speakers do not (Miyawaki, Strange, Verbrugge, Liberman, Jenkins and Fujimura, 1975). For another example, not only do Spanish and American English speakers place their category boundaries at different points along the VOT continuum (Abramson and Lisker, 1973; Williams, 1977), but also Spanish-English bilinguals can be induced to shift their boundaries by a shift in language set within a single test (Elman, Diehl and Buchwald, 1977). Not unrelated, perhaps, is the recent demonstration by Ganong (1978) that listeners have a bias for words over nonwords: offered a continuum of which one end is a word (e.g., bag) and the other not (e.g., pag), they shift their normal boundary away from the word, thus increasing the number of words they hear.

Presumably there are limits to this sort of thing. With adequate synthesis, the range of uncertainty must be limited and we may still use synthetic continua to assess "the auditory tolerance of phonological categories" (Brady and Darwin, 1978, p. 1556) -- precisely the use for which they were first designed over twenty-five years ago.

Feature or property detectors

The demonstration that listeners can be trained to hear a supposedly categorical continuum noncategorically undercuts the original evidence for acoustic feature, or property, detectors in speech perception, namely, categorical perception itself. Moreover, it throws into doubt the interpretation of a substantial body of work on selective adaptation of speech sounds that has appeared in the past five years.



The series began with a paper by Eimas and Corbit (1973). They asked listeners to categorize members of a synthetic voice onset time (VOT) continuum (Lisker and Abramson, 1964) and demonstrated that the perceptual boundary between voiced and voiceless categories along that continuum was shifted by repeated exposure to (that is, adaptation with) either of the endpoint stimuli: there was a decrease in the frequency with which stimuli close to the original boundary were assigned to the adapted category and a consequent shift of the boundary toward the adapting stimulus. Since the effect could be obtained on a labial VOT continuum after adaptation with a syllable drawn from an alveolar VOT continuum, and vice versa, adaptation was clearly neither of the syllable as a whole nor of the unanalyzed phoneme, but of a feature within the syllable. Eimas and Corbit therefore termed the adaptation "selective" and attributed their results to the fatigue of specialized detectors and to the relative "sensitization" of opponent detectors. Subsequent studies replicated the results for VOT and extended them to other feature oppositions, such as place and manner of articulation. These studies have been reviewed by Cooper (1975), Ades (1976), and Eimas and Miller (1978).

Unfortunately, there are many grounds for doubting the opponent detector model. First, as already remarked, is the demonstration that listeners can be trained to discriminate at least some speech continua within categories. Second, the model lacks behavioral or neurological motivation. For, while the facts of additive color mixture make an opponent detector account of after-effects entirely plausible, the facts of laryngeal timing or spectral scatter at stop consonant onset certainly do not. Third, the hypothesis is rendered implausible by dozens of reports of contextual effects: adaptation of consonantal features is apparently specific to following vowel, to syllable position, to syllable structure (Hall and Blumstein, 1978) and even to fundamental frequency (Ades, 1977). As Simon and Studdert-Kennedy (1978) remark, "...the theoretical utility of selectively tuned feature detectors goes down as the number of contexts to which they must be tuned goes up." Moreover, the degree of adaptation varies quite generally with the acoustic distance between adaptor and test syllables, an effect typical of psychophysical contrast studies. In fact, Simon and Studdert-Kennedy (1978), drawing on their own work and that of Sawusch (1977), marshal evidence to show that selective adaptation along speech continua reflects a combination of peripheral auditory fatigue and central auditory contrast. They do not deny that selective adaptation has possible fruitful use in isolating functional channels of analysis. But if their argument is correct, we now have no evidence at all for specialized detector mechanisms tuned to the acoustic correlates of abstract linguistic

features.

#### Scaling studies and feature interactions

This conclusion sits nicely with the results of many studies in which phoneme confusions or similarity judgments have been used to characterize the psychological representation of speech sounds. Although results vary widely with experimental method (van den Broecke, 1976), these studies typically find that vowels (e.g., Terbeek, 1977) and consonants (e.g., Singh, Woods and Becker, 1972) fall readily into low-confusion/high-similarity groups isomorphic with some standard phonological feature set. However, as Goldstein (1977) has pointed out, relations within these feature groups are usually not random. Rather, the psychological space is structured in such a way as to suggest a continuous auditory representation within feature groups. Presumably, since the continuous auditory representation derives from an acoustic structure shaped by articulation, we could describe an analogous articulatory space by scaling articulatory errors. It was Goldstein's (1977) insight to hypothesize that the variance common to the auditory and articulatory spaces would then prove to be categorical. His study -- too complicated for summary here -- largely supported that hypothesis. We may fairly conclude that our models of perception should allow for continuous auditory and articulatory representations from which categories can only be derived by some abstract metric common to both.

The idea that speech sounds (perhaps unsegmented syllables) may be internally represented in a continuous auditory space (at some point before classification) is compatible with the repeated finding of interaction between features during perceptual processing (e.g., Sawusch and Pisoni, 1974; Miller, 1977). There is, in fact, no good reason to refer to these auditory processes as "featural" at all (Parker, 1977). Repp (1977) and Oden and Massaro (1978), for example, have already proposed specific models of integration based on a continuous spatial representation.

#### Steps toward an auditory-articulatory space

The view of speech perception that seems to be emerging from the studies we have reviewed is of an active, continuous process. We turn now to several studies of perceptual integration across the syllable which seem to call for just such an interpretation.

Perhaps the most familiar example is provided by voicing cues for stops in initial position. The concept of voice onset time (VOT) originally offered an articulatory account of how a range of disparate and incommensurable acoustic cues (including, as it happens, the interval between release burst and the onset of voicing) comes to signal the voiced-voiceless distinction.



In fact, as Abramson (1977) has recently reminded us, VOT is itself simply a special case of the laryngeal timing mechanisms by which voicing distinctions are, in general, implemented.

To illustrate the underlying articulatory rationale, consider the suggestion by Stevens and Klatt (1974) that the duration of the first formant voiced transition might be a more potent cue than VOT itself. The motivation for the proposal seems to have been to coordinate the voicing cue with Stevens' hypothesized cues to place of articulation (rapid spectral scatter), and perhaps to avoid saddling the infant with a delicate timing mechanism. As it happens, Simon and Fourcin (1978) have shown that English speaking children do not learn to use the F1 cue until they are five years old, while French-speaking children never use it at all. In any event, careful analysis by Lisker (1975) and by Summerfield and Haggard (1977) has shown that the principal first formant cue is not transition duration, but frequency at onset: the higher the frequency, the less likely is a sound to be judged voiced. Listeners apparently take a high first formant onset as a cue that the mouth was relatively wide open (and release therefore well past) when voicing began.

A less familiar set of cues to another distinction has recently been studied by Repp, Liberman, Eccardt and Pesetsky (1978). They recorded the utterance: "Did anybody see the gray ship?" Then, by varying the durations of fricative noise at the onset of ship and of the silent interval between gray and ship, they explored the conditions under which the utterance was heard as ending with "gray chip," "great ship" or "great chip." Among their results was the finding that whether or not a syllable final stop was heard (gray vs. great) depended not only on the duration of the silence, but also on the duration of the noise following the silence. Just such an equivalence between a spectral property and silence emerges from an analysis of the trading relation between silence and formant transition in the cues for the medial [p] of [split] (Liberman and Pisoni, 1977). How are we to rationalize such an equivalence? Repp, et al. (1978) point out that neither a single feature detector nor a set of feature detectors, integrated by some higher level decision mechanism (as proposed by Massaro and Cohen, 1977), nor, it would seem, any purely auditory principle can explain why such phenomenologically diverse cues can be traded off and integrated into a unitary percept.

As a final example, consider a positively daedalian series of experiments by Bailey and Summerfield (1978). They explored the conditions under which a particular voiceless stop ([p], [t] or [k]) is perceived if a silence is introduced between [s] and a following vowel. Whether a stop is heard at all depends, of course, on the duration of the silence, but the effect of that

duration itself depends on the onset frequency of F1, while the perceived place of articulation depends on the duration of the closure, on spectral properties at the offset of [s] and on the relation between those properties and the following vowel (cf. Dorman, et al., 1977). Bailey and Summerfield suggest that, "...given sufficiently precise stimulus control, perceptual sensitivity could be demonstrated to every difference between two articulations" (p. 55) (cf. Haggard, 1977). Again, the problem is to understand the principles by which such heterogeneous collections of spectral and temporal cues are combined into a percept. What rationalizes their integration?

The answer, explicitly proposed by the authors of these several studies, is that the cues are held together by their origin in the integral, articulatory gesture. We should be absolutely clear that this is not a form of motor theory. Rather, it is a description of what the perceptual system appears to do. The system follows the moment-to-moment acoustic flow, apprehending an auditory "motion picture", as it were, of the articulation, in a manner totally analogous to that by which the visual system might follow the optic flow to apprehend the articulation by reflected light rather than by radiated sound. (cf. Fowler, submitted; Studdert-Kennedy, 1977).

#### Reading lips and reading spectrograms

The argument is clarified, and developed, in a recent study of lip reading by Summerfield (unpublished Ms). Subjects were asked to write down a series of sentences spoken over an audio system, but simultaneously masked by the talker's own voice reading another text. There were three conditions of interest to the present discussion: (1) audio alone; (2) audio with full video of the speaker's face; (3) audio with a video display of the speaker's lips. Without any training, naive subjects scored 23%, 65% and 54% correct, respectively. In a second experiment, Summerfield analyzed errors made against deliberately conflicting video. He found, as did McGurk and McDonald (1976), that subjects frequently made judgments reflecting a compound between the auditory and visual information. Summerfield (as also Haggard, 1977) points out that such instantaneous interplay between modalities seems to require a common metric by which the two streams of information can be combined. (The problem, incidentally, is quite general and may apply to any sound-producing visual event.)

It is instructive to compare the ease with which naive subjects used the visual display of face or lips with the obvious difficulty experienced by even the most skilled spectrogram reader. Cole, Rudnick, Reddy and Zue (1978) report a systematic study of subject VZ who has been studying acoustic phonetics for more than seven years and has logged some 2000-2500

hours reading spectrograms -- perhaps as many hours as a child of two years has spent listening to speech. Despite the fact that VZ is free to use the ample context of vision (rather than the narrow window of audition) and that he reports conscious, acoustic-phonetic interpretation of visual context at least 18% of the time; despite the fact that he came to the spectrograms knowing that their visual segments were not isomorphic with phonetic segments (a crucial piece of knowledge that cannot be derived from the spectrograms themselves); despite the fact that, in the hours devoted to spectrograms, he could probably have learned to read several foreign languages with fair proficiency, VZ now transcribes spectrograms at a rate some 20 to 40 times real time (Cole, personal communication).

One is not surprised. There are, after all, biological constraints on learning (see Hinde and Stevenson-Hinde, 1973): pigeons learn more readily to peck plastic keys for grain and to jump to avoid shock than vice versa. The visual display of talking lips and face is natural and its code is known to every speaker of a natural language, as the code of a spectrographic display is not. Watching its mother's face and listening to her speak, the infant learns to perceive articulation directly, whether by light or by sound.

#### Extracting information from the syllable

The primary unit of perception is evidently the unsegmented syllable (the rhythmic unit of nursery rhymes), and there is ample evidence for perceptual interaction between its components (see Studdert-Kennedy, 1976, for a review). For a recent example, Hasegawa and Daniloff (1976) synthesized two fricative continua, /s/ - /ʃ/, before two different vowels, /i/ and /u/, and found a significant shift in the phoneme boundary as a function of following vowel. Kunisaki and Fujisaki (1977) developed the finding by showing that contextual dependency in perception corrects for a mirror-image contextual dependency in production: just as the frequencies of fricative poles and zeros are lower before /u/ than before /a/, so, in perception, the frequencies of the poles and zeros at the synthetic boundary between /s/ and /ʃ/ are higher before /a/ than before /u/. These results mesh neatly with our earlier conclusion that consonantal onset is judged as part of a dynamic, temporal pattern.

Just such a process has recently been shown to play an important role also in vowel perception. Strange, Jenkins and Edman (1978) recorded tokens of /b/-vowel-/b/ syllables with ten different medial vowels, spoken by several speakers. They edited out the steady-state syllable nuclei (50% to 65% of the entire syllable, depending on the vowel) and presented various fragments

of the syllables for identification. The results varied with both speaker and vowel, but overall, for three speakers of the same dialect as the listeners, error rates on the original syllables, on the syllables without their centers ("silent centers") and on the isolated centers were 4%, 10% and 18% respectively. The error rates for either the initial or the final transitions alone were approximately 60%. Evidently, the dynamic sweep of the spectral information and its temporal distribution across the syllable was the principal source of listener information in identifying these vowels, even when that portion usually said to characterize a vowel (namely, its steady state) was completely missing.

Results such as these return us to the segmentation issue. Clearly, there was little basis for peripheral segmentation in these syllables. In fact, one is tempted to suppose that listeners recognized syllables (Massaro, 1975) or perhaps "diphones" (Klatt, in press a) rather than phonemes. Mermelstein (1978) reports a subtle experiment that speaks to this issue. He varied the duration and first formant frequency of the steady-state nucleus of synthetic syllables to yield /bed/, /bæd/, /bet/, /bæt/. Notice that exactly the same acoustic information (namely, duration of the steady-state nucleus) controls both vowel and final consonant decision. Accordingly, if subjects are asked to determine duration boundaries for both consonant voicing and vowel quality as a function of F1 frequency, and if the boundaries prove to be correlated, then we can conclude that listeners made a single -- presumably syllabic -- decision. However, if the boundary values prove independent, we can conclude that listeners recognized phonemes rather than syllables and that they made two phonetic decisions on the basis of a single piece of acoustic information. This was, in fact, the outcome. If this is the normal mode of speech perception, it would seem that, even if syllabic segmentation is peripheral (cf. Myers, et al., 1975), phonemic segmentation may be a central process consequent upon classification. Usually, this process is facilitated by auditory contrast within the syllable (cf. Bondarko, 1969).

#### Continuous speech

We come, finally, full circle to continuous speech with its prosody, syntax and "real world" constraints. Here, the main question is whether the perceptual processes we have been discussing up to this point have any bearing at all. Is it possible, for example, that, given the contextual aids of prosody, syntax, semantics, the listener needs no more than the "auditory contour" of a word (Nooteboom et al., 1976; cf. Morton and Long, 1976) or perhaps a few "invariant features" (Cole and Jakimik, in press) to gain access to his lexicon?

I have no space for a full discussion of this issue (a beginning is made by Liberman and Studdert-Kennedy, 1977). But a good place to start is with a paper by Shockey and Reddy (1975) who studied speech recognition in the absence of phonological and all other higher order constraints. They recorded some fifty short utterances, spoken by native speakers of eleven different languages and presented them to four trained phoneticians for transcription. The transcriptions were then compared with a "target" description, determined from native speakers and spectral analysis. The average "correct" score for the four transcribers was 56% and their average agreement 50%. Comparable scores for transcription of a familiar language, without contextual or syntactic constraints, would be roughly 90% -- the level reached by the three transcribers of Cole, et al., (1978) in their spectrogram reading study, cited above, and, moreover, a level close to that of VZ himself when reading spectrograms. The difference of roughly 40% is evidently due to the transcribers' knowledge of the phonology of the language being transcribed.

The point of this example is that the main difference between listening to continuous speech in a familiar language and to isolated words in a foreign one may not be in the syntax, semantics or real world constraints so much as in the phonology. This is a simplification, since phonology and syntax are not independent. But it serves to emphasize that phonology makes linguistic communication possible by setting limits on how a speaker is permitted to articulate and what a listener can expect to hear (Liberman and Studdert-Kennedy, 1977). The problem of how the listener extracts and combines information from the signal to arrive at a unitary percept is, of course, exactly the same for continuous speech as for isolated words.

The function of the other higher order constraints -- syntax, context, semantics -- is facilitative. They serve to delimit the sampling space from which the listener's percepts may be drawn. This is well illustrated by several experiments of Cole and Jakimik (in press), using the ingenious "listening for mispronunciations" (LM) technique, devised by Cole (1973). Subjects are asked to listen to a recorded story into which mispronunciations have been systematically introduced. Their accuracy and speed of detection is then measured as a function of different variables. Mispronunciations prove to be more rapidly reported for high than for low transitional probability words (cf. Morton and Long, 1976), for words appropriate to a theme than for words inappropriate, for words implied by previous statements than for words not implied, and so on. Presumably the more rapid reports reflect the varied ways in which thresholds for words are lowered by contextual factors. Of course, the fact that listeners recover the words at all means that they can do so

without a full phonetic analysis. But this should not, in my opinion, be taken to mean that they can do so without any phonetic analysis at all.

By far the fullest and most careful account of the interactive processes of word recognition in continuous speech is offered by Marslen-Wilson (1975, 1978). His experimental procedure also involves mispronunciations, but the subjects' task is to shadow the text as rapidly as possible. Marslen-Wilson examines the effects of context on the frequency of fluent restorations. These restorations are often so fast that the shadower begins to say the correct word (e.g., "company") before the second syllable of the mispronounced word (e.g., "compsiny") has begun (cf. Kozhevnikov and Chistovich, 1965). Since such restorations only occur when the disrupted word is syntactically and semantically apt, it is evident that these higher order factors have facilitated recovery of the correct word. However, they cannot do so in the absence of all phonetic information. It is reassuring to read as the conclusion of a lengthy and subtle discussion of these matters: "...word-recognition in continuous speech is fundamentally data-driven, in the specific sense that the original selection of word-candidates is based on the acoustic-phonetic properties of the initial segment of the incoming word" (Marslen-Wilson and Welsh, 1978, p. 60). Perhaps all these years of studying CV syllables have not been wasted after all.

#### References

- Abercrombie, D. (1967): Elements of General Phonetics. Chicago: Aldine.
- Abramson, A. and L. Lisker (1973): "Voice timing perception in Spanish word-initial stops", J. of Phonetics 1, 1-8.
- Ades, A.E. (1976): "Adapting the property detectors for speech perception. In Wales, R.J. and Walker, E. New Approaches to Language Mechanisms, Amsterdam: North Holland.
- Ades, A.E. (1977): "Source assignment and feature extraction in speech", J. Exp. Psych.: Hum. Perc. and Perf., 3, 673-685.
- Ainsworth, W.A. (1972): "Duration as a cue in the recognition of synthetic vowels", J. Acoust. Soc. Amer., 51, 648-651.
- Allen, G. (1972): "The location of rhythmic stress beats in English: An experimental study", Language and Speech, 15, 72-100.
- Aslin, R.N. and D.B. Pisoni (1978): "Some Developmental Processes in Speech Perception", Paper presented at NICHD Conference, "Child Phonology: Perception, Production and Deviation", Bethesda, MD, May 28-31, 1978.
- Bailey, P.J. and Q. Summerfield. (1978): "Some observations on the perception of [s] + stop clusters. Haskins Laboratories Status Report, SR-53(2), 25-60.
- Blumstein, S.E. and K.N. Stevens (in press) "Acoustic Invariance in Speech Production", J. Acoust. Soc. Amer.
- Bondarko, L.V. (1969): "The syllable structure of speech and distinctive features of phonemes", Phonetica, 20, 1-40.

- Brady, S.A. and C.J. Darwin (1978): "Range effect in the perception of voicing", J. Acoust. Soc. Amer. 63, 1556-1558.
- Carney, A.E., G.P. Widin, and Viermeister, N.F. (1977): "Noncategorical Perception of Stop Consonants Differing in VOT", J. Acoust. Soc. Amer. 62, 961-970.
- Chistovich, L.A., N.A. Fyodorova, D.M. Lissenko, and M.G. Zhukova (1975): "Auditory segmentation of acoustic flow and its possible role in speech processing", In Fant, G. and Tatham, M.A.A. (eds.) Auditory Analysis and the Perception of Speech. New York: Academic Press, 221-232.
- Cohen, A. and S.G. Nootboom (eds.) (1975): Structure and Process in Speech Perception, New York: Springer-Verlag.
- Cole, R.A. (1973): "Listening for mispronunciations: A measure of what we hear during speech", Perception and Psychophysics 1, 153-156.
- Cole, R.A. and J. Jakimik (in press): "Understanding speech: How words are heard", In Underwood, G. (ed.) Strategies of information processing, New York: Academic Press.
- Cole, R.A. and B. Scott (1973): "Perception of temporal order in speech: The role of vowel transitions", Canadian J. Psych. 27, 441-449.
- Cole, R.A., A. Rudnicky, R. Reddy and V.W. Zue (1978): "Speech as patterns on paper", In Cole, R.A. (ed.) Perception and Production of Fluent Speech, New Jersey: Erlbaum Associates.
- Collier, R. and J. 't Hart (1975): "The role of intonation in speech perception", In Cohen, A. and Nootboom, S.G. (eds.), Structure and Process in Speech Perception, New York: Springer-Verlag, 107-123.
- Cooper, W.E. (1975): "Selective adaptation of speech", In Restle, F., Shiffrin, R.M., Castellan, N.J., Lindman, H., and Pisoni, D.B. (eds.), Cognitive Theory, New Jersey: Erlbaum Associates.
- Cooper, W.E. (1976): "Syntactic control of timing in speech production: A study of complement clauses", J. Phonetics 4, 151-171.
- Cooper, F.S., P.C. Delattre, A.M. Liberman, J.M. Borst, and L.J. Gerstman (1952): "Some experiments on the perception of synthetic speech sounds", J. Acoust. Soc. Amer. 24, 597-606.
- Cutler, A. (1976): "Phoneme-monitoring reaction time as a function of preceding intonation contour", Perception and Psychophysics 20, 55-60.
- Cutler, A. and D.J. Foss (1977): "On the role of sentence stress in sentence processing", Language and Speech 20, 1-10.
- Cutting, J.E. and B.S. Rosner (1974): "Categories and boundaries in speech and music", Perception and Psychophysics 16, 564-570.
- Darwin, C.J. (1975): "On the dynamic use of prosody in speech perception", In Cohen, A. and Nootboom, S.G. (eds.), Structure and Process in Speech Perception, New York: Springer-Verlag.
- Darwin, C.J. and C.E. Bethell-Fox (1977): "Pitch continuity and speech source attribution", J. Exp. Psych.: Human Perc. and Perf. 3, 665-672.
- Dorman, M.F., J.E. Cutting, and L. Raphael (1975): "Perception of temporal order in vowel sequences with and without formant transitions", J. Exp. Psych.: Human Perc. and Perf. 1, 121-129.
- Dorman, M.F., M. Studdert-Kennedy, and L.J. Raphael (1977): "Stop Consonant Recognition: Release Bursts and Formant Transitions as Functionally Equivalent, Context-Dependent Cues", Perception and Psychophysics 22, 109-122.
- Eimas, P.D. and J.D. Corbit (1973): "Selective adaptation of linguistic feature detectors", Cogn. Psych. 4, 99-109.
- Eimas, P.D. and J.L. Miller (1978): "Effects of selective adaptation on the perception of speech and visual patterns: Evidence for feature detectors", In Walk, R.D. and Pick, H.L., Jr. (eds.) Perception and Experience, New York: Plenum.
- Elman, J.L., R.L. Diehl, and S.E. Buchwald (1977): "Perceptual switching in bilinguals", J. Acoust. Soc. Amer. 62, 971-974.
- Fischer-Jørgensen, E. (1972): "Tape cutting experiments with Danish stop consonants in initial position", Annual Report VII, Institute of Phonetics, University of Copenhagen, Copenhagen, Denmark.
- Fowler, C.A. (Ms. submitted for publication): "Perceptual Centers in Speech Production and Perception", Perception and Psychophysics.
- Fujisaki, H., K. Nakamura and T. Imoto (1975): "Auditory perception of duration of speech and non-speech stimuli", In Fant, G. and Tatham, M.A.A. (eds.) Auditory Analysis and Perception of Speech, New York: Academic Press, 197-220.
- Ganong, F. (1978): "A Word Advantage in Phoneme Boundary Experiments", J. Acoust. Soc. Amer. 63, 520(A).
- Goldstein, L. (1977): "Categorical features in speech perception and production", To appear in: Fromkin, V. (ed.) Proceedings of the Workshop on Slips of the Tongue and Ear, Vienna, (Also, University of California at Los Angeles), Working Papers in Phonetics, 39.
- Haggard, M.P. (1977): "Do we want a theory of speech perception?", Paper presented to the Research Conference on Speech-Processing Aids for the Deaf, Gallaudet College, Washington, D.C., May 23-26.
- Hall, L.L. and S.E. Blumstein (1978): "The effect of syllabic stress and syllabic organization on the identification of speech sounds", Perception and Psychophysics 24, 137-144.
- 't Hart, J. and A. Cohen (1973): "Intonation by rule: A perceptual quest", J. Phonetics 1, 309-327.
- Hasegawa, A. and R.G. Daniloff (1976): "Effects of vowel context upon labeling the /s/ - /ʃ/ continuum", J. Acoust. Soc. Amer. 59, 525(A).
- Hinde, R.A. and J. Stevenson-Hinde (1973): Constraints on Learning, New York: Academic Press.
- Klatt, D.H. (in press, a): "Speech Perception: A Model of Acoustic Phonetic Analysis and Lexical Access", In Cole, R.A. (ed.) Perception and Production of Fluent Speech, New Jersey: Erlbaum Associates.
- Klatt, D.H. (in press, b): "Analysis and Synthesis of English /b,d,g/", In Lindblom, B.E.F. and Ohman, S.E.G. (eds.) Frontiers of Communication Research, New York: Academic Press.
- Klatt, D.H. (1977): "Review of the ARPA Speech Understanding Project", J. Acoust. Soc. Amer. 62, 1345-1366.
- Klatt, D.H. (1976): "Linguistic Uses of Segmental Duration in English: Acoustic and Perceptual Evidence", J. Acoust. Soc. Amer. 59, 1208-1221.
- Klatt, D.H. and W.E. Cooper (1975): "Perception of Segment Duration in Sentence Context", In Cohen, A. and Nootboom, S.G. (eds.) Structure and Process in Speech Perception, New York: Springer-Verlag, 69-89.

- Kozhevnikov, V.A. and L.A. Chistovich (1965): *Rech 'artikuliatsiia i vospriiatie*, Moscow-Leningrad. Transl. as *Speech articulation and perception*, Washington, D.C. Clearinghouse for Federal Scientific and Technical Information, JPRS 30-543.
- Kuhn, G.M. (1975): "On the front cavity resonance and its possible role in speech perception", *J. Acoust. Soc. Amer.* 58, 428-433.
- Kunisaki, O. and H. Fujisaki (1977): "On the influence of context upon perception of voiceless fricative consonants", *Annual Bulletin RILP* (University of Tokyo), 11, 85-91.
- Lehiste, I. (1977): "Isochrony reconsidered", *J. Phonetics* 5, 253-264.
- Liberman, A.M. and D.B. Pisoni (1977): "Evidence for a special speech-perceiving subsystem in the human", In Bullock, T.H. (ed.) *Recognition of Complex Acoustic Signals*, Berlin: Dahlem Konferenzen, 59-76.
- Liberman, A.M. and M. Studdert-Kennedy (1977): "Phonetic Perception" In Held, R., Leibowitz, H. and Teuber, H.-L., *Handbook of Sensory Physiology*, Vol. VIII, Heidelberg: Springer-Verlag.
- Liberman, A.M., K.S. Harris, H.S. Hoffman, and B.C. Griffith (1957): "The discrimination of speech sounds within and across phoneme boundaries", *J. Exp. Psychol.* 53, 358-368.
- Lindblom, B. and K. Rapp (1973): "Some temporal regularities of spoken Swedish", *Papers from the Institute of Linguistics*, University of Stockholm, 21, 1-59.
- Lisker, L. (1975): "Is it VOT or a first-formant transition detector?", *J. Acoust. Soc. Amer.* 57, 1547-1551.
- Lisker, L. and A. Abramson (1964): "A cross-language study of voicing in initial stops: Acoustical measurements", *Word* 20, 384-422.
- Marslen-Wilson, W.D. (1975): "Sentence perception as an interactive parallel process", *Science* 189, 226-228.
- Marslen-Wilson, W.D. and A. Welsh (1978): "Processing interactions and lexical access during word recognition in continuous speech", *Cogn. Psych.* 10, 29-63.
- Massaro, D.W. (1975): "Preperceptual images, processing time, and perceptual units in speech perception", In Massaro, D.W. (ed.) *Understanding Language*, New York: Academic Press.
- Massaro, D.W. and M.M. Cohen (1977): "Voice onset time and fundamental frequency as cues to the /zi/ - /si/ distinction", *Perception and Psychophysics* 22, 373-382.
- Martin, J.G. (1972): "Rhythmic (hierarchical) versus serial structure in speech and other behavior", *Psych. Rev.* 79, 487-509.
- McGurk, H. and J. McDonald (1976): "Hearing lips and seeing voices", *Nature* 264, 746-748.
- Mermelstein, P. (1975): "A phonetic context-controlled strategy for segmentation and phonetic labeling of speech", *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-23, 79-82.
- Mermelstein, P. (1978): "On the relationship between vowel and consonant identification when cued by the same acoustic information", *Perception and Psychophysics* 23, 331-336.
- Miller, J.L. (1977): "Nonindependence of feature processing in initial consonants", *J. Speech and Hearing Research* 20, 519-528.
- Miller, J.D., C.C. Wier, R. Pastore, W.J. Kelly and R.J. Dooling (1976): "Discrimination and Labeling of Noise-Buzz Sequences with Varying Noise-Lead Times: An Example of Categorical Perception", *J. Acoust. Soc. Amer.* 60, 410-417.
- Miyawaki, K., W. Strange, R. Verbrugge, A.M. Liberman, J.J. Jenkins, and O. Fujimura (1975): "An Effect of Linguistic Experience: The Discrimination of [r] and [l] by Native Speakers of Japanese and English", *Perception and Psychophysics* 18, 331-340.
- Morton, J. and J. Long (1976): "Effect of Word Transition Probability on Phoneme Identification", *J. Verbal Learning and Verbal Behavior* 15, 43-51.
- Morton, J., S. Marcus, and C. Frankish (1976): "Perceptual centers (P-centers)" *Psych. Rev.* 83, 405-408.
- Myers, T.F., M.G. Zhukova, L.A. Chistovich, and V.N. Mushnikov (1975): "Auditory segmentation and the method of dichotic stimulation", In Fant, G. and Tatham, M.A.A. (eds.), *Auditory Analysis and the Perception of Speech*, New York: Academic Press, 243-274.
- Nooteboom, S.G., J.P.L. Brokx, and J.J. de Rooij (1976): "Contributions of prosody to speech perception", Institute for Perception Research Preprint, Eindhoven, Netherlands (To be published in Levelt, W.J.M. and Flores d'Arcais, G.B., *Studies in Language Perception*, New York: Wiley).
- Oden, G.C. and D.W. Massaro (1978): "Integration of Featural Information in Speech Perception", *Psych. Rev.* 85, 172-191.
- Öhman, S.E.G. (1975): "What is it that we perceive when we perceive speech?", In Cohen, A. and Nooteboom, S.E.G. (eds.) *Structure and Process in Speech Perception*, New York: Springer-Verlag, 36-48.
- Parker, F. (1977): "Distinctive features and acoustic cues", *J. Acoust. Soc. Amer.* 62, 1051-1054.
- Pastore, R.E., W.A. Ahroon, K.J. Baffrito, C. Friedman, J.S. Puleo, and E.A. Fink (1977): "Common factor model of categorical perception", *J. Exp. Psych.: Human Perc. and Perf.* 3, 686-696.
- Pisoni, D.B. (1977): "Identification and Discrimination of the Relative Onset of Two Component Tones: Implications for the Perception of Voicing in Stops", *J. Acoust. Soc. Amer.* 61, 1352-1361.
- Pisoni, D.B. and J.H. Lazarus (1974): "Categorical and non-categorical modes of speech perception along the voicing continuum", *J. Acoust. Soc. Amer.* 55, 328-333.
- Repp, B.H. (1977): "Dichotic competition of speech sounds: The role of acoustic stimulus structure", *J. Exp. Psych.: Human Perc. and Perf.* 3, 37-50.
- Repp, B.H., A.M. Liberman, T. Eccardt, and D. Pesetsky (1978): "Perceptual Integration of Cues for Stop, Fricative and Affricate Manner", *Haskins Laboratories Status Report on Speech Research*, SR-53(2), 61-83.
- Samuel, A.G. (1977): "The effect of discrimination training on speech perception: Noncategorical perception", *Perception and Psychophysics* 22, 321-330.
- Sawusch, J.R. (1977): "Peripheral and central processing in speech perception", *J. Acoust. Soc. Amer.* 62, 738-750.
- Sawusch, J.R. and D.B. Pisoni (1974): "On the identification of place and voicing features in synthetic stop consonants", *J. Phonetics* 2, 181-194.

- Shields, J.L., A. McHugh, and J.G. Martin (1974): "Reaction time to phoneme targets as a function of rhythmic cues in continuous speech", J. Exp. Psych. 102, 250-255.
- Shockey, L. and R. Reddy (1975): "Quantitative analysis of speech perception", In Fant, G. (ed.) Proceedings of the Speech Communication Seminar (Stockholm, Sweden), New York: Wiley.
- Simon, C. and Fourcin, A.J. (1978): "Cross-language study of speech-pattern learning", J. Acoust. Soc. Amer. 63, 925-935.
- Simon, H.J. and M. Studdert-Kennedy (1978): "Selective Anchoring and Adaptation of Phonetic and Nonphonetic Continua", J. Acoust. Soc. Amer. 64.
- Singh, S. (1978): "Distinctive features: A measurement of consonant perception", In Singh, S. (ed.) Measurement procedures in speech, hearing and language, Baltimore: University Park Press, 93-155.
- Singh, S., D.R. Woods, and G.M. Becker (1972): "Perceptual structure of 22 prevocalic English consonants", J. Acoust. Soc. Amer. 52, 1698-1713.
- Stevens, K.N. (1975): "The potential role of property detectors in the perception of consonants", In Gant, G. and Tatham, M.A.A. (eds.), Auditory Analysis and Perception of Speech, New York: Academic Press, 303-330.
- Stevens, K.N. and D.H. Klatt (1974): "Role of formant transitions in the voiced-voiceless distinction for stops", J. Acoust. Soc. Amer. 55, 653-659.
- Strange, W. and J.J. Jenkins (1978): "The Role of Linguistic Experience in the Perception of Speech", In Pick, H.L., Jr. and Walk, R.D. (eds.), Perception and Experience, New York: Plenum Publishing Corp.
- Strange, W., J.J. Jenkins, and T.R. Edman (1978): "Dynamic information specifies vowel identity", J. Acoust. Soc. Amer. 63(A).
- Strange, W., R. Verbrugge, D.P. Shankweiler and T.R. Edman (1976): "Consonant environment specifies vowel identity", J. Acoust. Soc. Amer. 60, 213-224.
- Studdert-Kennedy, M. (1976): "Speech perception", In Lass, N.J. (ed.), Contemporary Issues in Experimental Phonetics, New York: Academic Press.
- Studdert-Kennedy, M. (1977): "Universals in phonetic structure and their role in linguistic communication", In Bullock, T.H. (ed.) Recognition of Complex Acoustic Signals, Berlin: Dahlem Konferenzen, 37-48.
- Studdert-Kennedy, M., A.M. Liberman, K.S. Harris, and F.S. Cooper (1970): "Motor theory of speech perception: A reply to Lane's critical review", Psych. Rev. 77, 234-249.
- Summerfield, Q. (1975): "How a full account of segmental perception depends on prosody and vice versa", In Cohen, A. and Nooteboom, S.G., Structure and Process in Speech Perception, New York: Springer-Verlag, 51-68.
- Summerfield, Q. (unpublished Ms.) "Use of visual information for phonetic perception".
- Summerfield, Q. and M. Haggard (1977): "On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants", J. Acoust. Soc. Amer. 62, 436-448.
- Svensson, S.G. (1974): "Prosody and Grammar in Speech Perception," University of Stockholm, MILUS 2.
- Terbeek, D. (1977): "Across-language multi-dimensional scaling of study of vowel perception", University of California at Los Angeles, Working Papers in Phonetics, 37.
- Van den Broecke, M.P.R. (1976): Hierarchies and Rank Orders in Distinctive Features, Netherlands: Van Gorcum-Assen.
- Williams, L. (1977): "The perception of stop consonant voicing by Spanish-English bilinguals", Perception and Psychophysics 21, 289-297.
- Zhukov, S.Ya., M.G. Zhukov, and L.A. Chistovich (1974): "Some new concepts in the auditory analysis of acoustic flow", Sov. Phys. Acoust. 20, 237-240. [Akust. Zh., 20, 386-392.]
- Zue, V.W. (1976): "Acoustic characteristics of stop consonants: A controlled study", Lincoln Laboratory Technical Report No. 523, M.I.T. Cambridge, MA.

#### Acknowledgements

I thank Alvin Liberman, David Pisoni, Terry Halwes and, especially, Bruno Repp for their careful comments on the text. Preparation of this paper was supported in part by NICHD Grant HD-01994.



## AUDITORY PROCESSING OF SPEECH

Ludmilla A. Chistovich, Pavlov Institute of Physiology of the Academy of Sciences of the USSR, Leningrad, USSR

I shall outline the approach (Chistovich, Ventsov, Granstrem et al., 1976) adopted by our group in studying auditory levels in speech perception. After reading Studdert-Kennedy's report, I realized that some explanation of our reasons should be presented.

When phoneticians describe the acoustic cues they usually refer to some "objects" or "events" seen on the dynamic spectrogram, such as gaps, transitions and so on. Studdert-Kennedy has presented very good evidence that the parameters of these events (for instance, the duration of a gap, the direction of a formant transition) as well as the temporal order of the events displayed over intervals of roughly syllabic length are utilized by the human being in the phonetic interpretation of the message. Unlike Studdert-Kennedy we were not able to suggest any procedure for automatic phonetic interpretation which would conform with known experimental data on speech perception without assuming the preliminary conversion of the speech signal into a flow of events. To make this clear I shall mention only one quite trivial problem - the problem of the measurement of duration. Duration is the interval of time between two events. This parameter does not exist at all if the events delimiting the interval are not specified. That seems sufficient to explain why our group became interested in the auditory bases for the detection of events.

Neurophysiological studies of the central auditory system have revealed a highly ordered tonotopic organization at each anatomical level, with several representations of the frequency scale at the same level. This suggests that the original peripheral excitation pattern is transformed into a number of versions, with the axes of the pattern remaining unchanged. Stimulus-response relations for the central auditory neurons hint at the extraction of irregularities in the pattern along both frequency and time axes. There are indications that the width of the window of processing increases both in time and in space (frequency range) at higher levels of the system.

This led us to believe that the detection of irregularities in the speech-induced excitation pattern might be an essential part

of signal processing, and that the psychoacoustic study of the detection of irregularities might be a good starting point. To see how models, based on psychoacoustical data, will react to real speech, one has to build them as working signal-processing systems. A functional model of the cochlea is a necessary instrument for this approach. Our group is exploiting a linear model of the cochlea built as a 128 channel analyzer (Goloveshkin et al., 1978). Parameters of the model have been adjusted according to tuning curves for auditory nerve fibers. Dynamic spectrograms of speech obtained from this model are somewhat different from the conventional dynamic spectrograms, but almost all the details believed to be important are preserved.

So far we have confined ourselves to the simple kinds of irregularities: irregularities of the envelope and irregularities of the spectrum shape of a steady-state stimulus.

#### Processing of the envelope

Slow changes in the stimulus envelope are perceived as loudness changes. Rapid solitary irregularities such as jumps, drops, small gaps, hills and valleys give rise to associations with consonants. Although subjects cannot indicate any particular consonant with certainty, they have no difficulty in deciding whether the consonant associations are the same for two stimuli and whether they are present or absent. This allows us to use the classical psychoacoustical approach and to measure the quantitative relations between parameters when their effects are equal. The data concerning relations between the parameters of envelope changes (Chistovich, Ventsov, Granstrem et al., 1976; Stoljarova and I. Chistovich, 1977; I. Chistovich, 1978) suggest processing close to band-pass filtering, with the center frequency of the envelope filter being around 25 Hz. Having assumed band-pass filtering, we had to decide whether it is sufficient to use a single filter in the model, with something like instantaneous loudness being the input signal, or whether it is necessary to use a number of filters, each processing information within a restricted frequency range. It was found that although the association of a small gap in a pure tone with [r] does exist over a wide frequency range, it disappears when the tones preceding and following the gap differ in frequency (Lesogor, 1977). The critical mistuning of the two tones appears to be close to the critical band well known in psycho-

acoustics. The masking of the jump in the envelope of one tone by a simultaneously presented second tone has also been studied (Lesogor et al., 1978). The tone with the jump was held constant in frequency, intensity and jump amplitude, while the frequency of the masker ( $F_m$ ) was varied and the minimal level ( $L_m$ ), necessary to make the "consonant" disappear, measured. The resulting  $L_m$  vs.  $F_m$  curve appeared to be very similar to the so-called "psychoacoustical tuning curves". These data indicate that the detection of envelope irregularities requires our model to be multichannel. The simplest solution is to place one envelope filter at the output of each channel of the "cochlea".

In the first version of our frequency selective model for processing stimulus envelopes (I. Chistovich, 1978), half-wave rectification with a memoryless compressive nonlinearity was used to simulate the mechanical-to-neural transformation in the cochlea. Two (positive and negative) thresholds were placed at the output of the envelope filter, their crossing resulting in onset- and offset-markers. Better agreement with psychoacoustical data was achieved when peripheral short-term adaptation was also incorporated in the model.

This multichannel model, including the "cochlea", has been built as an analog system (Kozhevnikov et al., 1978). The model is good at detecting the rapid spectral and intensity changes in speech signals, and could be used for the automatic segmentation of speech. Although it is blind to the formants in a steady-state stimulus, it succeeds in tracking formant transitions. The model is not yet satisfactory from the psychoacoustical point of view. It cannot reproduce the above mentioned frequency selective effects in jump and gap detection, which seem to require that some spatial (interchannel) interaction must be incorporated in the model.

A serious problem concerns the combining of markers over the frequency scale. Data on perception of amplitude irregularities on the widely spaced components of a complex stimulus (Rodionov et al., 1976; Lesogor and Chistovich, 1978) indicate some kind of summation over a wide frequency range. The temporal threshold for jump detection (minimal interval between the stimulus onset and the jump) appeared to be equal to the threshold of nonsimultaneity for the onsets of two tones (Kozhevnikova, 1978). This also points to



summation. The threshold was found to be insensitive to "selective adaptation" (Ogorodnikova, 1978).

Summation of the markers would be useful for locating exactly the moment of change, but it will lose information about the frequency region where the change occurs. So far we have failed to find any evidence that the subject is able to pick up the frequency component which is the carrier of the amplitude jump or the gap. At the same time we have found that the subject "knows" the stimulus spectrum shape at the moment when the irregularity occurs (Zhukov et al., 1974; Zhukov and Lissenko, 1974). When a small gap was moved along a [iu] stimulus (F<sub>1</sub> - steady-state, F<sub>2</sub> - time-varying), subjects were able to locate the point corresponding to the shift from [iru] to [igu]. Gaps with different durations were adjusted by subjects in such a way that the end of the gap always coincided with the same value of F<sub>2</sub>. Subjects could do this just as easily when the time-varying F<sub>2</sub> was presented to one ear while the gap (in the stimulus with steady-state formants) was presented to the other ear.

Segments of the signal between onset and offset markers cannot be regarded as phonetic elements at this stage of processing. Temporal rules are used by the subject in accepting (or rejecting) the vowel-like segment as a vowel - the element of rhythmic pattern. These rules are based on segment duration as well as on the duration of the onset-to-onset interval (between one segment and its successor) and on the offset-to-offset interval (between one segment and its predecessor) (Chistovich, Ventsov, Granstrem et al., 1976).

#### Spectrum shape processing

Two-formant stimuli with widely spaced formants are convenient for measuring the formant peak detection threshold, since the criterion of a shift in vowel quality can be used. The fact that the threshold depends on both the formant spacing and the steepness of the spectrum slope (Mushnikov and Chistovich, 1971; Chistovich et al., in press) suggests a process such as spatial differentiation of the excitation pattern. Stimuli with spectral peaks just below threshold and just above threshold have been used to adjust the parameters of a lateral inhibition model processing the output pattern of the "cochlea". The weighting function (spatial window) appeared to be quite narrow. The output of the model to

a natural steady-state vowel is a spatial pattern with a number of peaks separated by zero-excitation intervals. To convert this pattern into a conventional formant description of the vowel, one has to identify its peaks with formants of the appropriate serial number and pick up the coordinate values (frequency position and amplitude) corresponding to the peaks. This procedure seems rather unrealistic from the point of view of the neurophysiology of hearing. The "center of gravity" effect (Delattre et al., 1952) indicates the spatial integration of a spectral pattern and suggests that the intermediate formant description of a stimulus might not be necessary for it to be identified as a vowel.

To test the "center of gravity" effect single-formant stimuli and two-formant stimuli with 350 Hz formant spacing (AI > A2 and AI < A2) have been used. Clear evidence for the effect was found in both the identification data and the matching data (Bedrov et al., 1978).

The "center of gravity" effect can be described in a qualitative way as  $F_1 < F^* < F_2$ , where  $F^*$  is the frequency of the single-formant stimulus most close in vowel quality to the two-formant stimulus. Maximal spacing of the formants in the two-formant stimuli and the range of the formant amplitude ratio delimiting the area of the existence of the effect have been measured (Chistovich et al., in press). The critical spacing appeared to be equal to 3.0 - 3.5 Bark and the amplitude ratio range could reach 40 dB. Experiments on two-formant to two-formant matching for stimuli with more-than-critical formant spacing indicate that in this case the formant amplitudes are of minor importance, provided both formant peaks are above threshold. Stimuli with quite different A<sub>1</sub>/A<sub>2</sub> values are most similar in vowel quality when their formant frequencies coincide. The data suggest a model with spectral peaks extracted at a lower level of processing and spatial integration at a higher level.

Our current attempts to simulate both the "center of gravity" effect and the unimportance of formant amplitudes with widely spaced formants apply a rather small set of spatial summators with overlapping summation intervals, each summator corresponding to one particular cardinal vowel. Assuming that the stimulus is described in terms of the distribution of the amount of excitation in the subset of excited summators, one is able, by using the model as an

instrument, to evaluate the similarity of two stimuli in vowel quality and to carry out the matching experiment. The set of cardinal vowels seems to be a better approximation to the inventory of vowels "known" by a Russian subject than the set of Russian phonemes. This follows from both the mimicking data (Avakjan, 1976) and the similarity scaling data (Kuznetsov, 1978). I should like to note that identifying summators with cardinal vowels is in fact one version of the template approach of which Studdert-Kennedy seems to disapprove.

There is no doubt that at least some of the parameters of the model must be made context-sensitive. A very strong adaptation-like effect was observed in the experiments on formant peak detection (Chistovich et al., in press). The nature of the effect is not yet analyzed.

#### Spectral cues in nonstationary vowels

The temporal parameters of spectrum shape processing are not yet known. It seemed useful to find out first what cues in the time-varying spectral shape pattern are important to the subject. Experiments with short two-formant vowel-like stimuli with linear and close to triangular (up-down and down-up) F<sub>2</sub> contours revealed three cues used in phonetic interpretation (Lublinskaja and Slepokurova, 1977; 1978). One cue corresponds to F<sub>2</sub> or the spectrum shape value at the "target" point: the extreme point of the triangular contour and the end-point of the linear contour. The second cue corresponds to the initial value of F<sub>2</sub> or of the spectrum shape. The third cue is the direction of the initial F<sub>2</sub>-transition. This last cue appeared to be effective only in a restricted frequency range since it only serves to differentiate between [ɹ] and [ʀ] and between [ø] or [œ] and [ɛ] or [ə]. The boundary in the direction of the F<sub>2</sub>-transition is somewhat displaced from the zero transition, the amount of the displacement being systematically different in different subjects. It would be very interesting to find out whether the transitions utilized in consonant perception are represented by different complex events (for instance, the transitions which occur not later than some critical interval from the onset marker) or whether they are the same as the transitions differentiating vowels.

In conclusion I would like to present one topic for discussion. We (our group) believe that the only way to describe human

speech perception is to describe not the perception itself but the artificial speech understanding system which is most compatible with the experimental data obtained in speech perception research. The main point is that artificial systems are based on many sources of scientific information, speech perception data being only one of these sources. If our point of view is accepted (I doubt that speech psychologists will agree with us), then it will be practical to direct experimental research to those problems which arise in automatic speech processing research.

Let us discuss some problems in automatic "phonetic processing"; they are most relevant to this meeting. The main problems concern the input parameters (representation of the signal at the input of the processor), the output representation and the rules and procedures of transformation. To specify the output one has to decide what kind of inventory (phonemes, allophones or something else) to use and how to represent the prosodic information. These problems are especially important from the point of view of simulating the higher levels of processing. Fortunately, research in this field does not really depend on exact knowledge of the lower levels of processing. In the case of the identification rules the situation is basically different. The rules are bound to depend strictly on the form of the signal representation: if you change the parameters extracted from the signal you must change the identification rules. It would be a good strategy to concentrate effort on the problems of auditory processing and on constructing automatic systems to simulate this processing. With these systems in hand it would be possible to approach the problem of rules by using both speech perception methods and the statistical methods applied in automatic speech recognition research.

#### References

- Avakjan, R.V. (1976): "Study of the perception of isolated vowels by Russian language users", *Fiziologia cheloveka* 2, 81-90.
- Bedrov, Ja. A., L.A. Chistovich and R.L. Sheikin (1978): "Frequency location of the "center of gravity" of the formants as the useful parameter in vowel perception", *Akust. Zh.* 24, 480-486.
- Chistovich, I.A. (1978): "A functional model for envelope processing in the frequency channel of the auditory system", *Fiziologia cheloveka* 4, 208-212.
- Chistovich, L.A., A.V. Ventsov, M.P. Granstrem, S.Ja. Zhukov, M.G. Zhukova, E.K. Karnickaja, V.A. Kozhevnikov, D.M. Lissenko, V.V. Lublinskaja, V.N. Mushnikov, N.A. Slepokurova, N.A. Fedorova, R.Haavel, I.A. Chistovich and V.S. Shupljakov (1976): Physiology of speech. Speech perception, Leningrad: Nauka.

- Chistovich, L.A., R.L. Sheikin and V.V. Lublinskaja (in press): "Centers of gravity" and spectral peaks as the determinants of vowel quality", in Frontiers of Speech Communication Research, B. Lindblom and S. Ohman (eds.), London: Academic Press.
- Delattre, P., A.M. Liberman, F.S. Cooper and Gerstman (1952): "An experimental study of the acoustic determinants of vowel color: observations on one- and two-formant vowels synthesized from spectrographic patterns", Word 8, 195-210.
- Goloveshkin, V.T., V.S. Shupljakov, L. Bastet and J.M. Dolmazon (1978): "Functional model of auditory spectral analyzer (linear version)", in Avtomaticheskoe raspoznavanie slukhovykh obrazov 10, 23-24, Tbilisi.
- Kozhevnikova, E.V. (1978): "Natural categorization of stimuli with different delay of amplitude jump", Fiziol. zh. SSSR 64, 1843-1849.
- Kozhevnikov, V.A., V.D. Rodionov, E.I. Stoljarova and I.A. Chistovich (1978): "Study and modelling of the auditory extraction of amplitude irregularities", in Avtomaticheskoe raspoznavanie slukhovykh obrazov 10, 37-38, Tbilisi.
- Kuznetsov, V.B. (1978): "Phonetic interpretation of steady-state vowels", in Avtomaticheskoe raspoznavanie slukhovykh obrazov 10, 97-98, Tbilisi.
- Lesogor, L.V. (1977): "Measurement of the admissible mistuning in r-gap perception", Fiziologia cheloveka 3, 85-87.
- Lesogor, L.V., V.D. Rodionov and L.A. Chistovich (1978): "Masking of the irregularity on the tone envelope by a tone of different frequency", Akust. zh. 24, 563-568.
- Lesogor, L.V. and L.A. Chistovich (1978): "Detection of consonant in two-component complex sounds and interpretation of stimulus as a sequence of elements", Fiziologia cheloveka 4, 213-219.
- Lublinskaja, V.V. and N.A. Slepokurova (1977): "Perception of vowel-like sounds with time-varying spectra", Fiziologia cheloveka 3, 77-84.
- Lublinskaja, V.V. and N.A. Slepokurova (1978): "Perception of synthetic vowels with initial and final second formant transition", in Avtomaticheskoe raspoznavanie slukhovykh obrazov 10, 99-100, Tbilisi.
- Mushnikov, V.N. and L.A. Chistovich (1971): "Auditory representation of the vowel. II. Detection of the second formant in the synthetic vowel", in Analiz rechevykh signalov chelovekom 11-19, Leningrad: Nauka.
- Ogorodnikova, E.A. (1978): "Investigation of 'selective adaptation' effect in simple nonspeech perception", Fiziol. zh. SSSR 64, 1831-1835.
- Rodionov, V.D., P. Carre and V.A. Kozhevnikov (1976): "Combining the information on short changes of the envelopes of the signals in different channels of the auditory system", Fiziologia cheloveka 2, 1021-1027.
- Stoljarova, E.I. and I.A. Chistovich (1977): "Frequency response of the model for auditory processing of envelope and the output threshold devices", Fiziologia cheloveka 3, 72-76.
- Zhukov, S.Ja. and D.M. Lissenko (1974): "Segmentation markers and their role in the interpretation of the formant contour", in Avtomaticheskoe raspoznavanie slukhovykh obrazov 8, Lvov.
- Zhukov, S.Ja., M.G. Zhukova and L.A. Chistovich (1974): "Some new concepts in the auditory analysis of acoustic flow", Akust. zh. 20, 386-392.

## SOME REMARKS ON RECENT ISSUES IN SPEECH PERCEPTION RESEARCH

Hiroya Fujisaki, University of Tokyo, Tokyo, Japan

I understand that the role of my contribution is to supplement Michael Studdert-Kennedy's comprehensive and impressive report. Therefore I will not try to give here an extensive review, but will state my personal remarks on some of the issues in recent studies of speech perception.

Categorical Perception of Speech and Non-speech Stimuli

A number of recent studies (Cutting and Rosner, 1974; Miller et al., 1976; Pisoni, 1977; Pastore et al., 1977) have confirmed the earlier assertion (Lane, 1965) that the categorical effect in discrimination measurements (I prefer the above expression to the conventional "categorical perception") is not specific to speech perception. As it has already been shown by a rigorous psychophysical account of the measurement procedure (Fujisaki, 1971), the apparent enhancement of discriminability across a category boundary (*not* the suppression of discriminability within categories) is an artifact that accrues from the subject's ability to categorize the test stimuli and to retain the results in the short-term memory, regardless of whether the stimuli are speech or non-speech. In other words, the categorical effect is a consequence of the single fact that the subject possesses or is provided with a stable threshold for categorical judgment of individual stimuli, but the process of discrimination is clearly sequential since the comparative judgment for discrimination is mediated by the results of categorical judgment. This simply indicates the inherent inability of our test procedures to dissociate the two types of judgment. But why are people so eager to look into, and to produce still new examples of, this phenomenon, when, after all, discriminability plays only a minor role in the actual speech communication? One of the interesting outcomes of these efforts may be the indication of perceptual similarity between the VOT continuum of stop consonants and some non-speech continua, suggesting that the perception of speech categories might be based on some simple psychoacoustic properties rather than on complex speech-specific properties. Generalization of this finding to other speech sound categories, however, requires careful investigations since there exist a number of acoustic continua on which phoneme categorization is not universal, but is more or less specific to individual languages.

### Levels of Processing and Selective Adaptation

There is little doubt that the identification of a particular segment of speech is a categorical judgment based on a number of acoustic properties (cues) detected from the continuous speech signal. For the sake of simplicity, we shall drop the issue of segmentation and defer the discussion of contextual effects to a later section. Conceptually, therefore, phoneme identification can be regarded as a two-stage process: property detection and decision. Neurophysiological evidences of signal processing in the visual cortex (e.g. Hubel and Wiesel, 1965), however, suggest that the detection of these properties is performed by a large number of neurons or neuron groups, arranged in a multi-level structure rather than a single-level structure; a set of primary properties being utilized for extracting a secondary property at the next level, and a set of secondary properties being further utilized for extracting a property of a still higher order at the next level, etc. Thus the conventional division of two levels (auditory vs. phonetic, or peripheral vs. central) may not be appropriate and the transition from the peripheral to the central processing may be more gradual than it is suggested by the terminology. It should also be noted that the extraction of individual properties need not be competitive, and the final decision is made after combination and temporal integration of the higher-order properties (Repp et al., 1978). The selective adaptation paradigm (Eimas and Corbit, 1973) is certainly a powerful tool to look into these mechanisms. Through systematic manipulation of the properties to be shared by the adaptors and the test stimuli as well as of the modes of stimulus presentation and response (e.g. Sawush, 1977a, 1977b), both structural and functional informations on these mechanisms have been accumulated. It is to be noted, however, that the adaptation is generally not restricted to one particular property detector nor to one particular level, and the resulting changes in a subject's response should be ascribed not only to changes in the sensitivity of the related property detectors but also to changes in the thresholds of categorical judgments both for phoneme identification and for stimulus rating. Further research on the elaboration of the paradigm, as well as its application to various speech sound categories other than the intensively studied voiced stops, would clearly lead to a deeper understanding of the processes of speech perception at least at the level of the phoneme.

### Speech Perception in Context

Although the selective adaptation paradigm is successful in studying the mechanisms of speech perception by creating a very special context, the results are not directly applicable to the process of speech perception in an ordinary context, where individual phonemes generally follow one after another and overlap in their articulatory realization to form a continuous acoustic string. Both articulatory and acoustic studies of monosyllables, as the smallest units of a phoneme sequence, reveal the mutual character of coarticulatory influences between the vowel as the syllable nucleus and the adjoining consonant(s). These coarticulatory changes are, however, compensated for by perception. For example, it is a well-known fact that the perception of voiced consonants is severely impaired if we take away the formant transitions and leave only the bursts (e.g. Dorman et al., 1977). Likewise, the perception of the syllable nucleus is incomplete when we take away the formant transitions to the adjoining consonant(s) and leave only the stationary portion (e.g. Fujimura and Ochiai, 1963; Strange et al., 1978). Thus the vowel and the consonant(s) within a syllable complement each other in perception. In more generalized connected speech, however, the coarticulatory influences extend over the syllable boundaries, and the perception of a vowel within a syllable is found to be incomplete if the syllable is taken out of its context and presented in isolation, but is restored when the syllable is presented with its immediately adjacent syllable(s) (Kawahara and Sakai, 1972). The perceptual mechanism of compensation for the coarticulatory variations of vowels has been investigated using synthetic disyllables of Japanese consisting of two vowels and non-speech stimuli with similar dynamic characteristics (Fujisaki and Sekimoto, 1975), indicating that the perception of a vowel in a dynamic context involves at least two distinct processes: extrapolation of incomplete formant transitions occurring both for speech and for non-speech, and short-term change of category boundaries occurring only for speech. Further investigation of the process of speech perception in the dynamic context is clearly necessary in order to elucidate the basis upon which the listener's knowledge of the language at the phonological, morphological, lexical, and syntactic levels, as well as the semantic and pragmatic information, is fully utilized in the understanding of spoken messages.

### The Roles of Prosody

Although prosody is not a well-defined concept, I consider it as a set of functions imposed upon a sequence of phonemes for the purpose of transmitting information concerning some linguistic units that are larger than the phoneme, such as word, sentence, and paragraph. Word prosody is almost synonymous with word accent (or intonation) and is used to transmit lexical information concerning homonyms. Sentence prosody consists of prominence, intonation, and rhythm, which are used to transmit or supplement both semantic and syntactic information of a sentence. Paragraph prosody (Lehiste, 1975; 1978), a relatively new concept, may be regarded as transmitting the structural information of a discourse. In addition to these major functions, prosody also contributes to facilitate segmental perception and to maintain the coherence of an utterance (Nootheboom et al., 1976), but I consider the latter functions to be rather subsidiary. These prosodic functions are realized mainly through the medium of suprasegmental features such as pitch, loudness and quantity (duration) of segments as well as of pauses, but may also be manifested by some segmental features such as phonemic quality of vowels (e.g. word accent in English). In spite of the importance of these functions in speech perception, comparatively little effort seems to have been spent in studying their perceptual effects. This may be firstly because of the lack or insufficiency of their formal descriptions, secondly because of the lack of analysis techniques to obtain quantitative acoustic formulations, and thirdly because of the increased difficulty in the preparation of synthetic speech stimuli of larger duration necessary for free and precise control of suprasegmental features. However, studies on perception of word accent and/or sentence intonation have recently been published on Japanese (Fujisaki and Sugito, 1976), on Dutch ('t Hart, 1976), on Danish (Thorsen, 1976), on Thai (Abramson, 1977), on Estonian (Eek, 1977), etc. The perceptual role of duration for expressing syntactic information has also been demonstrated for English (Lehiste et al., 1976). Perceptual reality of isochrony has been discussed and demonstrated using natural and synthetic speech (Lehiste, 1977; Higuchi and Fujisaki, 1978; Sato, 1978). On the other hand, perceptual roles of acoustic correlates of paragraph prosody, such as the peak in the fundamental frequency, pre-boundary lengthening, and pause duration, have been investigated using natural and spectrally-inverted utterances (Lehiste, 1975; 1978).

### Development and Impairments of Speech Perception

While the main interest of phoneticians may reside in the understanding of speech perception by an adult with normal hearing and language abilities, much could be learned from the study of developmental processes in young children (e.g. Fourcin, 1978), as aptly pointed out by Studdert-Kennedy. Studies of speech perception in hearing-impaired children (Fourcin et al., 1978; Waldman et al., 1978) are indispensable for finding systematic methods of training and for designing useful aids. Specially designed rhyme tests using natural utterances (Risberg, 1976) or identification tests using synthetic stimuli (Yokkaichi and Fujisaki, 1978) are useful for efficient collection of data on segmental perception, while the ability of identifying intonation contours can be tested by using natural utterances (Risberg and Agelfors, 1978). Furthermore, the perceptual ability of children and adults with language comprehension impairments can be tested by synthetic stimuli with various temporal characteristics (Tallal et al., 1976; Tallal and Newcombe, 1978), allowing one to locate the processing of rapid transitions at the dominant hemisphere. Studies of speech perception in its developmental stages as well as in the pathological cases can thus shed light on the process of speech perception in normal adults and can also lead to a better use of our knowledge for the alleviation of the impairments.

### References

- Abramson, A. (1977): "The phonetic plausibility of the segmentation of tones in Thai phonology," Haskins Labs. Status Rept. on Speech Research SR-53(1), 73-77.
- Cutting, J. and B.S. Rosner (1974): "Categories and boundaries in speech and music," Perc.Psych. 16, 564-570.
- Dorman, M.F., M. Studdert-Kennedy, and L.J. Raphael (1977): "Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues," Perc. Psych. 22, 109-112.
- Eek, A. (1977): "Experiments on the perception of some word series in Estonian," Estonian Papers in Phonetics 1977, 7-32.
- Eimas, P.D. and J.D. Corbit (1973): "Selective adaptation of linguistic feature detectors," Cogn.Psych. 4, 99-109.
- Fourcin, A.J. (1978): "Acoustic patterns and speech acquisition," Speech and Hearing, University College London 3, 143-172.
- Fourcin, A.J., S. Evershed, J. Fisher, A. King, A. Parker, and R. Wright (1978): "Perception and production of speech patterns by hearing-impaired children," Speech and Hearing, University College London 3, 173-204.

- Fujimura, O. and K. Ochiai (1963): "Vowel identification and phonetic contexts," JASA 35, 1889 (A).
- Fujisaki, H. and T. Kawashima (1971): "A model of the mechanisms for speech perception — Quantitative analysis of categorical effects in discrimination —," Ann.Rept.Engg.Res.Inst., Fac.Engg., Univ.Tokyo 30, 59-68. Also to appear as "On the modes and mechanisms of speech perception — Analysis and interpretation of categorical effects in discrimination," in Frontiers of Speech Communication Research, B. Lindblom and S. Ohman (eds.), London: Academic Press.
- Fujisaki, H. and S. Sekimoto (1975): "Perception of time-varying resonance frequencies in speech and non-speech stimuli," in Structure and Process in Speech Perception, A. Cohen and S.G. Nootboom (eds.), 269-280, Berlin/Heidelberg/New York: Springer-Verlag.
- Fujisaki, H. and M. Sugito (1976): "Acoustic and perceptual analysis of two-mora word accent types in the Osaka dialect," Ann. Bull. RILP, Univ. Tokyo, 10, 157-172.
- 't Hart, J. (1976): "Psychoacoustic backgrounds of pitch contour stylisation," IPO Annual Progress Report 11, 11-19.
- Higuchi, N., H. Fujisaki, and S. Sekimoto (1978): "Production and perception of segmental durations in spoken Japanese," JASA 64, Supplement No.1, S113 (A).
- Hubel D.H. and T.N. Wiesel (1965): "Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat," J.Neurophysiol. 28, 229-289.
- Kuwahara, H. and H. Sakai (1972): "Perception of vowels and C-V syllables segmented from connected speech", Journal of the Acoustical Society of Japan 28, 225-234.
- Lane, H. (1965): "The motor theory of speech perception: A critical review," Psych.Rev. 72, 275-309.
- Lehiste, I. (1975): "The phonetic structure of paragraphs," in Structure and Process in Speech Perception, A. Cohen and S.G. Nootboom (eds.), 195-203, Berlin/Heidelberg/New York: Springer-Verlag.
- Lehiste, I., J.P. Olive, and L.A. Streeter (1976): "Role of duration in disambiguating syntactically ambiguous sentences," JASA 60, 1199-1202.
- Lehiste, I. (1977): "Isochrony reconsidered," JPh 5, 253-263.
- Lehiste, I. (1978): "Temporal organization and prosody — perceptual aspects," JASA 64, Supplement No.1, S112 (A).
- Miller, J.D., C.C. Weir, R.E. Pastore, W.J. Kelly, and R.J. Dooling (1976): "Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception," JASA 60, 410-417.
- Nootboom, S.G., J.P.L. Brokx, and J.J. de Rooij (1976): "Contributions of prosody to speech perception," IPO Annual Progress Report 11, 34-54.
- Pastore, R.E., W.A. Ahroon, K.J. Baffuto, C. Friedman, J.S. Puelo, and E.A. Fink (1977): "Common-factor model of categorical perception," J.Exp.Psych. 3, 686-696.
- Pisoni, D. (1977): "Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops," JASA 61, 1352-1361.
- Repp, B.H., A.M. Liberman, T. Eccardt, and D. Pesetsky (1978): "Perceptual integration of acoustic cues for stop, fricative, and affricate manner," Haskins Labs. Status Rept. on Speech Research SR-53(2), 61-83.
- Risberg, A. (1976): "Diagnostic rhyme test for speech audiometry with severely hard of hearing and profoundly deaf children," STL-QPSR 2-3/1976, 40-58.
- Risberg, A. and E. Agelfors (1978): "On the identification of intonation contours by hearing-impaired listeners," STL-QPSR 2-3/1978, 51-61.
- Sato, H. (1978): "Temporal characteristics of spoken words in Japanese," JASA 64, Supplement No.1, S113 (A).
- Sawush, J.R. (1977a): "Peripheral and central processes in selective adaptation of place of articulation in stop consonants," JASA 62, 738-750.
- Sawush, J.R. (1977b): "Processing of place information in stop consonants," Perc.Psych. 22, 417-426.
- Strange, W., J.J. Jenkins, and T.R. Edman (1978): "Dynamic information specifies vowel identity," JASA 63, Supplement No.1, S5 (A).
- Tallal, P., R. Stark, and B. Curtiss (1976): "The relation between speech perception impairment and speech production impairment in children with developmental dysphasia," Brain and Language 3, 305-317.
- Tallal, P. and F. Newcombe (1978): "Impairment of auditory perception and language comprehension in dysphasia," Brain and Language 5, 13-24.
- Thorsen, N. (1976): "An acoustical investigation of Danish intonation: Preliminary results," Annual Report of the Institute of Phonetics, University of Copenhagen 10, 85-148.
- Waldman, F.R., S. Singh, and M.E. Hayden (1978): "A comparison of speech-sound production and discrimination in children with functional articulation disorders," L&S 21, 205-220.
- Yokkaichi, A. and H. Fujisaki (1978): "Identification of synthetic speech stimuli by hearing-impaired subjects," JASA 64, Supplement No.1, S19 (A).



## P h o n o l o g y

Report:	HANS BASBØLL	103
Co-Report:	STEPHEN R. ANDERSON	
	Notes on the development of phonological theory	133
Co-Report:	JOAN B. COOPER	
	Formal and substantive approaches to phonology	143

## PHONOLOGY

Hans Basbøll, Nordic Institute, Odense University, Denmark

The report to follow is a personal evaluation of some trends in phonology which have been more or less dominant since the last congress, as well as an overview of a subjective selection of individual contributions to phonology during the same period.<sup>1</sup> It is the reporter's hope that the most important lacunae of the text and bibliography will be filled by the co-reporters and the audience, and that the personal form and content of the report will provoke rather than prevent discussion.

A few words should be said about what will at most be covered in passing in this report. A number of the most central issues in current phonological debate have been selected as topics for semi-plenary symposia: Phonetic universals in phonological systems and their explanation, The psychological reality of phonological descriptions, Acquisition of the phonological system of the mother tongue, Social factors in sound change, and The relation between sentence prosody and word prosody (stress and tone). Consequently, these subjects will only be mentioned briefly or not at all in the present paper, and we shall not devote much attention either to the topic of the syllable in phonological theory, which will be treated in a working group.

Whereas section 1 is devoted to some general points characteristic of post SPE models of generative phonology (in the broad sense), there is no section of the present paper covering exclusively non-generative types of phonology. Some new theoretical developments of general interest within such theories will be mentioned in sections 2 and 3, however. It is outside the aim of this report to cover phonological descriptions of individual languages which are not intended to be contributions to phonological theory as well. The above principle of demarcation for this report is of course by no means to be taken as implying that phono-

1) I am indebted to the following people who have made a number of useful comments (both concerning content and style) on an incomplete version of the manuscript: Laurie and Winifred Bauer, Niels Davidsen-Nielsen, John Dienhart, Stig Eliasson, Eli Fischer-Jørgensen, Leif Kvistgård Jakobsen, Per Linell, and Jørgen Rischel. Unwisely, I have only followed some of their suggestions, and the responsibility for all flaws of the paper is, of course, mine alone.



logical analyses of individual languages are without scientific merit or interest. On the contrary, such descriptions are the fundament of our discipline, and a theorizing without a solid foundation in careful phonological and phonetic analyses, using field work and not merely reinterpreting the findings of others, is doomed to be an image (beautiful though it may be) with feet of clay.

## 1. Some trends and developments in generative phonology

### 1.1 Is there a school of generative phonology?

In his interesting overview with the characteristic title "Phonology since generative phonology", which in fact almost exclusively covers "the field of natural phonology", Bailey writes (1976, 5): "The writer's European experience convinces him that many linguists outside America believe that the newer phonology is just another development within generative phonology. For the most part, this is certainly not true. In fact, most natural phonologists rebelled as early as 1968 against generative phonology, often against the entire framework...". This point of view is not uncommon, but is nevertheless only true with certain modifications. In addition to the historical continuity of persons (and partly of institutions in the widest sense) natural phonology - and one could include other theoretical developments as well - is a continuation of standard generative phonology in the following two respects: there still seems to be a common basis of argumentation or, to put it in a simple-minded fashion, the various scholars speak closely related languages and understand each other reasonably well; and finally, but importantly, there is a crucial common core of theoretical references, both as concerns published work and, more fatally, semi-published or privately circulated papers. This may be less of a linguistic problem and more related to the field of the sociology of science, but I think that the notion of a linguistic or other scientific "school" belongs to the latter sphere too. All this is not to deny the existence of fundamental disagreements between standard generative and, say, natural phonology, but only to make clear why I think it is reasonable to speak of a generative "school" of phonology in the following. Opponents of the present view should compare standard generative phonology to both natural phonology (to see the similarity) and to, say, the "functional" trend of Martinet, or to stratificational phono-

logy, or to Soviet phonology (except Šaumjan), just to see the difference. Not only is generative phonology (including the field of natural phonology in the broad sense of Bailey 1976<sup>2</sup>) a "school" in this sense<sup>3</sup>, it is the dominant school as shown by the many references to it in phonological work of other theoretical persuasions, whereas reference within the generative school to competing theoretical views outside the school is frequently scarce, to say the least (except concerning sources of data or such predecessors to the school as Bloomfieldian linguistics). That insiders are in general less willing than outsiders to identify their own scientific context as a school is not a surprising state of affairs.<sup>4</sup>

### 1.2 Two main trends in generative phonology

I believe it is possible to discern two main directions within the evolution of later generative phonology, although all such rough categorizations must, of course, be taken with at least a grain of salt. The trends, or attitudes, which I have in mind might be termed "substance based" and "formal", respectively. To clarify this proposed distinction and its relation to standard generative phonology, let us briefly consider what may happen if a certain formulation of a phonological rule, e.g. from SPE, is taken really seriously.

One can focus upon the correspondence between the rule as stated, or even one part of the rule, and observables, i.e. give the rule a direct psychological interpretation or interpret it in

2) Bailey distinguishes four different trends within "Natural phonology": NP in the original form of e.g. Stampe, Drachman and Dressler (cf. now Donegan and Stampe, forthcoming); "Natural generative phonology" or "Concrete phonology I", worked out by Vennemann, Hooper and Rudes; "phonetology" or "dynamic phonology", i.e. Bailey's own trend; and finally, the phonology of e.g. Wang and Chen (which Bailey terms "Concrete phonology II"). One should be careful not to overlook the differences between these four types of "natural phonology", e.g. with respect to the distinction between "formal" and "substance based" generative phonology (section 1.2).

3) Of course, many levels of subschools seem to exist in this sense, both dominant and dominated. Consider, e.g., the fact that Hooper's textbook (1976) makes no reference to Bailey's work.

4) This observation also applies to non-generative linguistics, of course. Consider, e.g., the common claim among Danish linguists that there never was such a thing as a "Copenhagen school" (cf. Fischer-Jørgensen 1975a, 114). In the sense used here, this is not quite true, but it may of course be correct in other respects (as argued in Fischer-Jørgensen, loc. cit.).

some other way which is directly comparable with certain facts (cf. section 2 about 'evidence'). Such a "substance based" attitude is hardly compatible with a very "abstract" interpretation of the competence:performance distinction: it probably presupposes that competence is a 'competence for performance' where the path between the two is direct, short and due to factors which are in principle known (e.g. from studies of memory limitations).

An elegant example of a critical piece of argumentation which merely interprets (in the above sense) the details of a number of rule formulations taken from SPE is Stampe 1973. John Ohala and others have tried to test such rule formulations directly. An increasing number of authors have come to realize, however, that it may be a better research strategy (further see the end of the present section) to try and identify natural processes from "external" but "real life" data (sound change, speech disturbances, fast speech phenomena, language acquisition, etc.) and to consider each rule, or sometimes even a block of several rules, a unity for that purpose (on different types of rules, cf. section 1.4). The focus of interest is thus no longer isolated rule formulations, or even their parts, but functionally defined processes.

The "formal" trend alluded to above takes the formalism seriously in a different way. It is not the psychological or other empirical interpretation of the single parts of the formalism which is in focus (apart from certain premature claims about what is "psychologically real"), but partly the formal ingredients of the system (such as 'constraints' and a number of alleged 'formal universals' supposed to be innate, all with the purpose of narrowing 'the class of possible grammars of a natural language'), partly the generative capacity of the system as a whole (in this case considered as a black box). This trend agrees with Chomsky's position as evidenced in, e.g., "Conditions on transformations" and "Reflections on language"<sup>5</sup>, and it is particularly well represented in the recent journal Linguistic Analysis (which according to its characteristic subtitle aims to cover studies in formal syntax,

5) Chomsky's earlier position, on the other hand, has clear connections with logical empiricism, as evidenced e.g. by the striking similarities between "Syntactic structures" (1957) and the introductory part of Carnap's "Logische Syntax der Sprache" (1934).

semantics and phonology). The strong Chomskyan position may be illustrated by a handful of quotations from Koster et al. (1978, 3-4): "Human cognitive behaviour involves the interaction of diverse cognitive structures. [...] A direct route to performance, use, process, and the like, seems ill-conceived, because it would involve the result of interacting factors that are themselves unknown. [...] The analysis of cognitive structures has to precede the study of the enormously intricate synthesis which we call behaviour [...] The kind of cognitive psychology we advocate therefore rejects the holistic study of behaviour as hopelessly premature." The ultimate goal is "to account for the language faculty, and hence for the linguistic theory (the theory of Universal Grammar), in terms of human biology." There are a number of epistemological and methodological problems in this attitude (cf., e.g., Derwing 1973), but it is an interesting and maybe surprising fact that this strong Chomskyan position seems to have had hardly any consequences for linguistic analyses and argumentation as compared with that of other formalists within the generative school (like Milner) who reject that their object of study is anything like the state of a mental organ: in fact, only linguistic evidence is accepted.<sup>6</sup> Regardless of whether or not adherents of what I have labelled the formal trend of generative phonology consider their discipline as being a branch of cognitive psychology (in the Chomskyan sense), the analyses and explicit argumentation are thus in general intra-linguistic, and evidence from psychological tests and the like is quite generally not considered (due to the con-

6) Milner has concisely formulated his position like this: "Les propositions de la linguistique sont falsifiables, mais ne le sont que sur la base d'une évidence tirée des langues elles-mêmes. Aucune falsification tirée de l'évidence psychologique (ou biologique, ou de quelque ordre non-linguistique que ce soit) n'est donc pour moi admissible. Ce qui me frappe, c'est que cette position est celle de tous (ou presque tous) les linguistes génératifs, y compris ceux qui admettent [que la réalité du langage et des langues soit de substance essentiellement psychologique, et qu'une réalité psychologique soit un état spécifiable d'un organe mental]. J'en conclus que [les deux propositions entre crochets/HB] ne jouent aucun rôle réel dans la construction de la théorie linguistique" (1978, 9).

ception of the competence:performance distinction).<sup>7</sup>

These two directions of evolution within the school of generative phonology have been distinguished and presented in this way mainly for expository reasons. Although most concrete phonologists belong to the "substance based" trend, there is also a certain amount of formality here; and even though e.g. the scholars around Koutsoudas are formalists in the sense used here, they in fact sometimes make use of substantive evidence. In short, the bifurcation presented here is based upon several elements which are logically and empirically distinct, and furthermore the "substance based": "formal" distinction is not strictly binary: the two terms mark the endpoints of a scale.<sup>8</sup> A version of this scale has sometimes been known as the "abstract:concrete"-opposition, crucial to all phonological theory and practice, and the first of the converging tendencies we shall consider below is precisely the non-abstractness of lexical representations.

### 1.3 Non-abstractness of lexical representations and the issue of directionality

Abstraction is an inescapable condition for all sorts of descriptions including scientific ones, i.e. some aspects of the object to be described must necessarily be disregarded in order to obtain a description. However, the notions of "abstraction" and

7) Per Linell (personal communication) interprets the distinction between the "substance based" and the "formal" trend like this: Phonologists belonging to the former trend aim at describing language specific rules of certain ("linguistic") aspects of the production and perception of speech, whereas the latter type conceive of significant phonological generalizations as pertaining to much more abstract ("cognitive" or "mental") principles, presupposing - rather arbitrarily - that intralinguistic methods can yield such "cognitive" results.

8) More than anything else, I think the two proposed trends differ with respect to research strategy: The "formal" phonologists consider the rules and notation as given for a certain purpose, thus drawing conclusions concerning the interaction of rules etc. from the notation (cf. the use of models in theoretical physics). The "substance based" phonologists, on the other hand, do not accept any proposed rules without recourse to data outside normal linguistic behaviour (cf. certain "empirical" types of psychology). The two trends thus differ with respect to their general confidence in the proposed formal systems of phonology. Both attitudes may per se be scientific, their difference lies mainly in what they consider fruitful lines of research in the present state of our phonological knowledge (cf. section 2).

"abstractness" play a more crucial role in phonology than in most other scientific disciplines (including linguistic ones), since one distinctive trait of phonology as compared to phonetics can be claimed to be one of abstractness, with the further proviso that what has disappeared as a result of the "abstracting away" or reduction (together with the linguistic and non-linguistic context, and so on) is the phonetic details.

The above remarks apply to both generative and structural phonology. And in fact there seem in principle to be two distinct ways of abstracting from phonetic details to phonological forms (for discussion, see e.g. Rischel 1974, 361-365)<sup>9</sup>: One can either remove more and more redundancy from the class of possible pronunciations - within the language norm in question, of course - of a given word form; or one can go backwards in the derivation, so to speak, within a rule component constructed to account for (morphological) relatedness between different word forms. Although both of these types of abstraction have been used in both structural and generative phonology, the emphasis laid in these two theories clearly differs: structural phonology favours the first type, generative phonology the latter. The notion of surface contrast, which is essential in many structuralist schools of phonology, is reasonably well defined<sup>10</sup> except for the possible identification of members of different inventories belonging to distinct positions in the chain. If one goes further toward abstract forms, however, it is hard to find non-arbitrary criteria for where to stop the abstraction, in structuralist as well as in generative types of phonology.

9) It might be added, however, that this should not be taken to imply that semantics or pragmatics is necessarily more abstract than phonetics, although this implication may be tempting to both phoneticians and generative linguists. I should rather say - from a European structuralist point of view - that phonology is an abstraction vis-a-vis phonetics, in much the same way as semantics is an abstraction vis-a-vis pragmatics.

10) Bailey's interpretation of the "traditional phoneme" (1976, 14f) does not seem quite fair to me, e.g. as regards the Prague school notion of the phoneme (including the concepts of 'neutralization' and 'archiphoneme', which have now been revived in natural phonology): "-merely a redundancy-free phone. What few (less than a dozen) predictions, trivial or non-trivial, can be wrung out of this now ancient artifact all seem to be wrong -- not least those involving linguistic change and psychological reality".

As mentioned in the previous status report on phonology (Fischer-Jørgensen 1975b), one main development in the early seventies within generative phonology was in the direction of more concrete analyses. This trend could be seen partly as a reaction against very abstract phonologies as exemplified by Schane 1968, SPE, and numerous works by Lightner.<sup>11</sup> The problem with these abstract analyses was, of course, that they were consistent applications of the basic principles of generative phonology, and at the same time it appeared intuitively evident to most phonologists that they were highly implausible candidates for being components of a grammar which purportedly should be psychologically real. The fundamental reason why Schane, Lightner and others could reasonably arrive at such abstract analyses is that there was no operational criterion for the degree or type of relatedness between two word forms which would decide when one should posit a common underlying form and rules to make the derivation work (cf. Rischel 1978); and the simplicity criteria in use favoured common base forms in cases where a number of 'apparently unrelated' word forms could be related with only modest cost of rule complication (the generalizations were presupposed to be 'linguistically significant' but this concept had not been operationally defined either; however, cf. now Hurford 1977). Until today, not very much progress has been made concerning the establishment of criteria for relatedness between word forms (but B. Derwing has initiated research in that area). Instead, a number of authors have taken another route to reduce the run-away abstraction which can be tolerated in standard generative phonology: to find explicit constraints on the abstractness of the analyses, either on the lexical representations, or on the rules or the way in which they interact (see the next section), or in a combination of these.

A number of authors (e.g. Vennemann, Linell, Hooper, Rudes) - some of them even with a markedly 'abstract' past - have proposed (more or less) similar criteria on lexical representations

-----  
 11) The position of Foley (1977) is quite isolated: He criticizes SPE-phonology (which he rebaptizes "transformational phonetics") for being much too concrete, and favours a very abstract, non-psychological phonology. His theoretical views are reminiscent of those of glossematics about immanence and substance-independent glossems. The present writer agrees that SPE argues too much from the notation, but I fail to see why one should exclude oneself from phonetic explanations, e.g. in the case of strength hierarchies (cf. section 3.2).

to the effect that these should correspond to surface forms in distinct pronunciations, but not necessarily with detailed phonetic specifications. Such a constraint gives rise to reasonable analyses, e.g. in Hooper's version (1976). It should be pointed out, however, that if the lexical representations are hypotheses about how speakers actually store their phonological information regarding individual lexical items, then they should in principle be falsifiable by "external" criteria (it is evident that analyses are not "psychologically real" just because they are concrete, cf. section 2). On the other hand, if the lexicon is seen as a collection of phenomena - in this case pertaining to pronunciation and perception - which are not predictable by rule, then the lexical representations cannot be considered hypotheses about anything outside the grammar itself (and thus empirical vacuity may result), since they will then be negatively defined by the notion 'rule', which in this context seems to mean any regularity that can be stated.

If the lexical representations are claimed to have some sort of psychological reality, it will of course be no argument against the anti-abstract proposals just mentioned that they are highly redundant (this would presuppose an additional premise to the effect that information is stored in the brain in the most economical (compact) way, whereas the amount of computation needed to derive the actual forms, as well as different forms of retrieval, are less 'costly' for the overall system). One may challenge the plausibility of such concrete lexical representations as proposed e.g. by Rudes 1976 (syllabified whole but phonetically incompletely specified word forms) in view of (1) the amount of fully productive (both semantically, morphosyntactically and phonologically) formation of words, particularly in languages like Eskimo, and (2) the human ability to syllabify sound chains according to rules, in slow-careful speech as well as in allegretto speech, etc. It must be remembered, however, that the possible psychological reality of the lexical representations is an empirical issue that should be subjected to rigorous testing, but this is only possible after a further clarification of the notion 'psychological reality' (cf. Linell, forthcoming).

A consequence of the postulation of more concrete lexical representations may be that phonological rules are divided into

more abstract ("pre-lexical") rules (morphophonological or the like), and more concrete ("post-lexical") rules (phonetic or the like). A division of phonology into two types of phonological rules, 'abstract' and 'concrete', by no means presupposes concrete lexical representations, however. This issue will be taken up in the next section.

Another conceivable constraint that would automatically reduce the abstractness of lexical representations is the claim that all phonological rules should be bidirectional, or inferable, or (directly) recoverable, i.e. that the underlying form should be inferred from the surface (different formulations of such a constraint are possible, and it may pertain to rules, representations, or both).<sup>12</sup> The True Generalization Condition as used in Hooper 1976 (which in a sense generalizes proposals of Stanley 1967) in fact is such a constraint. Even authors who do not favour such a strong constraint have made use of the notion of recoverability, e.g. Gussman (1976). Eliasson in a number of interesting papers explores the notions of 'unidirectionality' and 'bidirectionality' in phonology, and he concludes that bidirectionality plays a much larger role than is usually ascribed to it in generative phonology.<sup>13</sup>

To end this section, let us briefly consider an apparently somewhat bizarre variation of generative phonology which nevertheless is not without virtues, viz. Leben and Robinson's "Upside-down phonology". Its basic idea is that the lexical representations are concrete surface forms (following Vennemann 1974), and that the whole machinery of e.g. SPE operates in the reverse of

12) It is evident that the formulation of the rules has an impact on the formulation of the lexical representations, and vice versa, and thus even strong constraints on only one of these factors may have very little over-all effect on the abstractness of the theory as a whole.

13) This renewed interest in bidirectionality is not only reminiscent of the bi-uniqueness criterion of Bloomfieldian phonemics, but also, e.g., of the stratificational classification of relations between levels in terms of neutralization, diversification, etc. As shown by Eliasson (e.g. forthcoming), there is clearly much insight to be gained from combining those structuralist viewpoints with the findings of generative phonology, and he explores e.g. various kinds of antiambiguity restrictions and historical restructurings which give substance to the notion of (partial) interconvertibility between levels.

the usual order (and thus ordering is necessarily extrinsic), not to determine the phonetic output (which was there in the first place), but to decide whether or not two forms are (morphologically) related (note that relatedness is thus not taken as something primary, as opposed to the rules). One undoes the phonological rules, one by one (and backwards, as stated), of the two word forms to be compared, and if they ever get alike during that process, then they are related. In fact, this restructuring of the standard generative model (into a parsing model) has a number of favourable effects, in particular concerning the notion 'analogy', as argued in the paper (although the criticisms of an overly concrete lexicon, of course, apply here too). One consequence of the model, when interpreted psychologically, is that surface similarity necessarily overrides paradigmatic regularity as an indicator of relatedness: e.g. obese-obesity (without vowel shift) are related by a 'shorter derivation', and thus - when the model gets a direct psychological interpretation - would seem more related (and, at any rate, not "exceptionally" related) than normal pairs like obscene-obscenity (with vowel shift). In that respect the "upside-down phonology" is not just a reinterpretation of the standard generative phonology.

#### 1.4 Functional variety of rules and their order of application

One major convergence in recent generative phonology (in the broad sense used here) is the division of phonological rules into at least two different main types: 'abstract' or 'morphophonemic' as against 'concrete' or 'phonetic' or 'allophonic' rules, sometimes called 'processes'. It should be pointed out from the outset that this dividing line falls within phonology as opposed to (pure) phonetics, i.e. it is not identical to the distinction between phonological rules proper and phonetic detail rules, e.g. in SPE, where the difference is that the distinctive features (at least the non-prosodic ones, in contrast to e.g. stress) are all binary in the former case, whereas they are 'scalar' in the latter (in a framework which permits non-binary distinctive features at the phonological level, cf. section 3 below, the characteristic trait of phonetic detail rules may reasonably be that the features vary continuously within a certain scale). If coarticulation effects (or even the fraction which may be language specific) should be accounted for by rule at all, it is certainly not by the type

of phonological rule used in generative phonology (and thus arguments like that of Bach 1968, repeated many times since then, to the effect that e.g. fronting of velars between front vowels is crucial evidence concerning the formal nature of rules and the simplicity metric seem misconceived from the outset). The new convergence described above is thus a dividing line within phonology itself, supported by a number of authors like Vennemann, Hooper, Bailey, Linell, Rischel, Drachman, Dressler and Koutsoudas. The dividing line is reminiscent of Kiparsky's (1973) distinction between neutralizing and allophonic rules. But in fact, a number of criteria which have been used, or may be used, do not classify rules in quite the same way (see Linell 1977 on a functionally based typology of phonological rules; also cf. Brasington 1976 and Dressler 1977a).<sup>14</sup> What is new is not only the distinction between an 'abstract' and a 'concrete' part of phonology, but also the emphasis on the latter.

In contradistinction to the authors mentioned above, Stephen R. Anderson (1975), while accepting the typological difference between 'morpholexical' and 'phonological' rules (with 'phonetic' rules as a third category, cf. above), claims - although not all of his arguments are wholly convincing to the present author - that they are interspersed (but he emphasizes that it may, in *casu*, be natural for a morphological rule to precede a phonological rule).

Some advances have been made in our understanding of the notion 'optional' rule (cf. Sanders 1977), partly from socio-linguistic investigations (e.g. by Labov and his associates). Also the influence of paralinguistic factors like speech tempo (cf. also Bolozky 1977) and style variation (as opposed to non-linguistic factors like sex, age and socio-economic group, layer or class), have come into the focus of attention, thanks not least to the work of Dressler and his colleagues. Due to such careful investigations, the psychological reality of word reduction phenomena has become apparent, as opposed to the realities described by many

-----  
 14) I should like to emphasize the following distinction which is not always observed in the literature: a phonotactic constraint (or condition) states which structures are permitted or prohibited, i.e. it is an intra-level notion; a phonotactically conditioned (or better, motivated) rule indicates only one means to obtain a certain phonotactic result and is thus an interlevel notion (the effect of such rules recalls what Kisseberth baptized 'conspiracies').

other phonological rules (cf. section 2).

Two of the criteria for the classification of rules mentioned in the present section, viz. morphophonological vs. phonetic and obligatory vs. optional, play a role in a certain general attitude to phonology, viz. one which adheres to the claim that all orderings in phonology can be predicted from a set of universal principles. A group of scholars around Koutsoudas (including Sanders, Noll, Iverson and Ringen) have investigated this hypothesis in a number of studies (starting with Koutsoudas et al., 1974), and a recent summary of the principles (Ringen 1976, 55f) lists the following: (1) The rules are scanned after each rule application to determine which rules are applicable to the new representation; (2) an obligatory rule must apply everywhere that its structural description is met unless some other principle predicts that it cannot apply; (3) rule A takes applicational precedence over rule B if the structural description of B properly includes the SD of A; (4) a derivation is completed when no more obligatory rules are applicable (and no more optional rules are opted for); (5) no rule can apply vacuously in any derivation (Ringen 1976, 57); and there is in addition a further principle (6) allowing consecutive and preventing nonconsecutive reapplications of a rule (a formulation is given in Ringen 1976, 62). It will be seen that principles (1) and (6) together with (4) and (5) define how rules are scanned and what counts as application and termination. (3) decides some cases where more than one rule is applicable, and further principles of this sort may be, and in fact have been proposed, e.g. that a morphophonemic rule takes precedence over an allophonic one (Koutsoudas 1977). Principle (2), finally, does not belong in any one category: it is partly a 'principle of precedence' (obligatory precedes optional), but, as pointed out by Ringen (*op. cit.*), that may be seen as a simple consequence of the meaning of the notion 'obligatory'; and the phrase 'unless some other principle predicts that it cannot apply' is a principle about the hierarchy among the principles themselves (viz. with respect to (3) here), i.e. a 'metaprinciple'.

The work just mentioned above clearly belongs to the "formal" trend of generative phonology (see section 1.2), and it is still controversial whether extrinsic ordering can be dispensed with within such a framework. Notice that this theory still allows



rules to be crucially and intricately ordered in derivations. In contradistinction to this, 'no ordering constraints' have also been proposed (e.g. by Vennemann) within natural phonology, i.e. within a "substance based" trend.

Two other directions of work within the "formal" trend of generative phonology concerning the application of rules deserve mentioning in the present context: One is the theory of local order proposed by S.R. Anderson (see e.g. 1974). A number of his original examples in favour of local order have been challenged recently (e.g. by Leben and Ringen), but I think one of my coreporters will give more information on that point if needed. The other is the phonological cycle in the SPE-sense (as opposed to e.g. that of Niles 1976, where the name 'cycle' is coupled to stylistic variation), which has been argued for by Brate and others. It is the opinion of the present writer that the phonological cycle in its original sense is unwarranted as a theoretical notion, cf. Rischel 1976.

In several papers (e.g. Basbøll 1976 with reference) I have generalised and applied McCawley's notion of rank so that every rule should apply with one of a small set of linearly ordered boundaries, including the syllabic one as its rank, i.e. with a string following by the boundary in question (or a stronger one) as its domain, and that a boundary should be allowed to occur properly only if it is the end of a rule. Furthermore, rules of a lower rank should apply only if a condition before rules of a higher rank, and this may be taken as a suggested further principle of precedence in the sense of a theory of 'no extrinsic order'.

References and articles in phonology  
In the book "Phonology: Linguistics as methodology" with the subtitle "On the rationality of non-empirical theories", Lass (1976) has given the following well known generative criterion for "methodology" as applied to "methodology" or "philosophical" (1976: 107): "The theory must be falsifiable". Lass con-

tinues with the following: "The theory must be falsifiable" (1976: 107). This is a well known generative criterion for "methodology" as applied to "methodology" or "philosophical" (1976: 107): "The theory must be falsifiable". Lass con-

cludes that most linguistic theories are not 'scientific' in this sense, in particular not Chomsky's theory of grammar although its creator repeatedly calls it so (this, of course, is not new, cf. well known criticisms by Botha, Derwing, Linell, and Itkonen): "If refutability is the hallmark of scientific theories, and if the empirical content of a theory is in direct proportion to its refutability, what are we to make of the majority of theoretical proposals in linguistics? [...] It is quite clear that many of them are infalsifiable for structural reasons. That is, they make claims for which no 'crucial experiment' or even reasonable testing procedure can be devised" (1976, 215). His own way out, still in agreement with Popper, is that theories which are neither demonstrable nor refutable may be respectable nevertheless if they are rationaly arguable: they try to solve certain problems, and it can be rationally discussed whether a certain solution is fruitful, simple, etc. in relation to the problem-situation in which it was devised. Demanding that linguistics should be empirical would mean, according to Lass (219ff), a shift of basic emphasis away from 'insight' in the normal linguistic sense, and restricting the field to those aspects which are capable (e.g. by means of 'rigorous experimentalism') of having empirical claims made about them.

It is the opinion of the present writer that Lass here goes too far in renouncing falsifiability (in favour of rational arguability) for most linguistic claims. The heart of the matter is, I think (cf. Spang-Hanssen 1959) that a scientific description should be prognostic, i.e. it should make predictions (which in principle could be refuted) about something outside the material on the basis of which it was constructed in the first place (this presupposes that the material is - in principle at least - considered open). This notion of prognosticity applies both to intra-, para- and extra-linguistic data. If this point of view is accepted then most linguistic statements, I think, are in principle refutable when new sets of data are considered (presupposed, of course, that the theoretical terms can be operationally defined). If the linguist is satisfied with rational argumentation and renounces refutation, he may be almost back in the sometimes futile discussions on 'simplicity', 'elegance', and so on, of several structuralist traditions. Although we must sometimes, e.g. in meta-



theoretical considerations, content ourselves with rational argumentation, a major goal of our discipline should - in my opinion - be to try to open as many areas of linguistics as possible to empirical investigation (i.e., to speak short-handedly, to potential refutation).

The previous status report (Fischer-Jørgensen 1975b) contained an evaluation of different types of external evidence and a rather detailed discussion of the notion 'psychological reality' in phonology. The program of research sketched there is tremendous, and clear results in these areas have, predictably, not been obtained in the meantime, so I shall limit myself to a reference to her report in this context.

Skousen (1975) investigates in detail a number of cases of "Substantive evidence in phonology". In contradistinction to Skousen, however, Dressler (1977b) has had divergent and incoherent results when using different modalities of external evidence, but this "is, hopefully, only true if one uses external evidence in a somewhat superficial way [...] Today higher standards must be set: first it must be argued why, in the first place, a particular modality of external evidence should be relevant for the specific problem in question, and what factors, warrants, and marginal conditions must be considered in order to ensure that the particular evidence really confirms what it should confirm, or can be explained in the same way as data from another modality. Here theory of science must come in ..." (Dressler 1977b, 224). Notice that these warnings by no means suggest that the linguist should limit himself to rational argumentation.

To close the section, a few words might be said about sound change. A number of recent investigations of chronological (and other) variation of language have increased our knowledge of the invariant aspects of human language as well. Examples of such studies are Chen and Wang 1975, Brink and Lund 1975, Lass 1976, and Bailey 1977a. A basic insight e.g. of the latter work is that natural processes should be kept strictly apart from non-natural (e.g. morphologized) rules which are spread by Creolization (the importance of sound change for the study of marking will be mentioned in section 3.3). A very promising comprehensive socio-linguistic investigation, viz. the Tyneside project (see Pellowe 1976) should also be mentioned.

### 3. Segments, features and marking

#### 3.1 The output of phonology: aspects of phonetic structuring

The question of the relation between phonetics and phonology is, of course, a vexed one (partly of a terminological nature, and both a normative and a descriptive one), and the most different opinions on this issue have had supporters in the past or the present, be it that they are identical, overlapping, properly included one in the other, non-overlapping, or in a relation of abstraction. A further possibility, in a "concrete" and "substance based" vein, is to use convention as the distinctive criterion such that 'phonology' should cover the language specific (conventional) and 'phonetics' the universal (biologically conditioned) aspects of sound structure. For the sake of clarity, we can put the question in the following form: Is phonology (in the broad sense used in this report) dependent on modern phonetic results, i.e. from physiological, acoustic or perceptual instrumental investigations? E.g., can it be the case that phonological theory has to be modified, or even radically changed, as the result of certain important new insights within phonetics? The question thus amounts to more than just asking whether phonology presupposes a certain basic phonetic knowledge (which probably no one would deny), and the answer depends on the phonologist who replies. What is interesting, however, is the fact that several new versions of phonology which build heavily on phonetic results have been propagated in print since the last congress. And I think it is fair to say that the understanding of the importance and even indispensability of phonetics in phonology is growing among phonologists. This evolution, which I for one appreciate, has been furthered by the work of phoneticians like Lindblom, Ladefoged, Fromkin, Lehiste and Ohala. As an example of this tendency a careful study on prenasalized consonants (with the revealing title "Phonetic analysis in phonological description") may be mentioned (Herbert 1977), in parallel to works on nasalization and palatalization by Chen and Mayerthaler, respectively. In the following, two more radical revisions of current phonological theory, viz. the auto-segmental and the non-segmental approach, will be considered in turn.

"Autosegmental phonology is", according to Goldsmith (1976, 23), "an attempt to provide a more adequate understanding of the phonetic side of the linguistic representation [...]; it suggests

that the phonetic representation is composed of a set of several simultaneous sequences of [segments, and, more concretely, it] is a theory of how the various components of the articulatory apparatus, i.e. the tongue, the lips, the larynx, the velum, are coordinated." It departs from the trivial but important phonetic observation that the speech chain cannot, phonetically, be sliced into a number of consecutive non-overlapping segments. Goldsmith proposes that certain features, mainly pitch but in some cases also nasality, should be treated on a level of their own (cf. the name 'auto-segmental'), and he examines the formal nature of the theory as well as a number of concrete cases (involving contour tones, tone stability, melody levels, floating tones, and automatic spreading of nasality) in support of the autosegmental view. His conclusion appears so sound to the present writer that it deserves to be quoted in part: "advances in phonological theory may start from an interest in low-level articulatory facts; [and] we do not begin our research with an understanding of the most elementary linguistic observables [...]. We should not restrict our attention to rules [...] at the risk of missing the very nature of the items involved." (1976, 67). As is immediately obvious even from the short summary above, the autosegmental approach shares a number of fundamental conceptions with the Firth school (or 'prosodic school'), although this historical aspect is not emphasized in Goldsmith 1976 (I think it would be a gain for our discipline if the work of our predecessors were taken into account more often than is the case today, cf. Fischer-Jørgensen 1975a). It should be added that Leben 1976 and Clements 1977 are interesting applications of the autosegmental approach to English intonation<sup>16</sup> and to vowel harmony, respectively.

An interesting and promising contribution to the theory of phonology since the last congress is T.D. Griffen's 'Non-segmental

---

16) Although the dividing line between phonetic and phonological models of intonation is by no means clear, a few important studies of English intonation with general linguistic implications might be mentioned in this report: Liberman 1975, Bailey 1977b, and Pellowe and Jones 1978.

phonology"<sup>17</sup> (see Griffen 1976 and 1977). Built upon recent advances in physiological phonetics (in particular the dynamic phonetic model of Mermelstein 1973), Griffen 1976 advances a phonological model in which the problems of segmentation in classical phonological theory, both structuralist and generative, are claimed to be overcome. He states - in agreement with e.g. Twaddell - that whereas the distinctive oppositions have observable correlates in phonetics, the segmental speech sound is nothing but a convenient fiction (partly due to the historical coincidence that writing when invented in the old world was alphabetical). Griffen "maintains a syllable in which the vowel is considered to be the articulatory base and consonants are constraints carried out on the vowel and concurrently with it" (1977, 375). This hierarchical notion of phonology which, as a matter of fact, reactualizes structuralist notions of hierarchy and dependency (cf. Rischel 1964 and Anderson and Jones 1974), is then applied to aspects of Modern Welsh. The new model has also been applied to a classical problem in phonology, viz. the relation between German [x] and [ç] (1977). It "eliminates the need for such allophones by attributing vowel characteristics to vowels and consonant characteristics to consonants" (ib.). Although this proposed explanation recalls prosodic analyses as well as Hockett (1955, 155-157), Griffen's proposal is interesting in itself because it follows from the so-called dynamic phonetic model. It is not improbable, however, that the conventional aspects of the distribution of German "ich" and "ach" are understated in Griffen's analysis. He claims that his model can describe the entire phonology by a simple hierarchical structure. To the present author, his analyses taken together seem rather convincing, but I find it a challenge for researchers with a major competence in modern phonetics to critically examine Griffen's model of non-segmental hierarchical phonology, and an important task for Griffen and others to develop and investigate this model

---

17) It should be noted that this use of the term "non-segmental" is not in agreement with that suggested by Chomsky and Halle where "non-segments" would mean "boundaries" (which, according to SPE 371, are units in the string with the feature [-segment]). This is, of course, no criticism of Griffen's use of the term, which is entirely reasonable and more immediately understandable than SPE's (whose conceptions of units and segments are, naturally, incompatible with Griffen's).

further. The main challenge to Griffen's theory is, as I see it, how it can be extended to deal adequately with a much wider range of phonological problems than have been covered within non-segmental hierarchical phonology until now.

### 3.2 The inventory and organization of features

It is probably an uncontroversial statement that some sort of distinctive features must have their place in a theory of phonology. A number of questions concerning such features which are anything but uncontroversial, however, will be briefly considered in turn (on marking, see section 3.3). I shall mainly build upon the work done in prolongation of Ladefoged 1971, which seems to me a more fruitful starting point for research in this area than e.g. SPE.

First of all, how should features be defined: articulatorily (cf. SPE), acoustically, perceptually, or in a combination (cf. Jakobson et al. 1952). The hybrid solution of Ladefoged (1971, 1975), Lindau (1975) and Williamson (1977) seems reasonable enough: they argue that the correlates of certain features are acoustically simple and articulatorily complex (e.g. "grave" - a feature which has also been argued for within an SPE-framework - and the basic features for vowel space according to Lindau 1975), and they should accordingly be defined acoustically. Other features should for a similar reason be defined articulatorily (e.g. "labial" - which has also been argued for within an SPE-framework - and "nasal"). This pragmatic view seems to the present writer to be reconcilable with the original Jakobsonian position, reemphasized by Henning Andersen, that the features are above all perceptual (although they will, in the present state of our knowledge, in general be better defined within other aspects of communication by sound-waves due to our lack of criteria for operational definitions within the realm of perception).

Another debated point is the question whether all features are binary. The strong binary position has never been convincingly argued for, in the opinion of the reporter. If the question of binarism is conceived of as an empirical one, the available evidence seems to suggest that some features are binary on the phonological level, e.g. nasality, and others multi-valued (with a small number of linearly ordered values), e.g. vowel height. The exact number of values of a feature is language specific within certain (biologically determined) limits. The preceding remarks apply to

a conception of phonology where the notion of surface contrast is in focus, but it still remains to be shown whether the question of binarism can be given any empirical content in much more abstract conceptions of phonology.

Concerning major class features, it is well known that SPE inherited the strange 'natural class' [h ? j w], defined as non-vocalic and non-consonantal sounds, from Jakobson et al. The drawbacks of this proposal have recently been discussed again (Lass 1976, 148-167). It is today generally accepted, I think, that the feature "vocalic" should be given up and the feature "syllabic" introduced instead (but cf. Andersen forthcoming). Problems arise, however, if "syllabic" is taken as a feature to be defined in a way which is parallel with other feature-definitions (cf. Ladefoged 1971, 94: "syllabic (correlates undefined)"). A better solution seems rather to be that 'syllabicity' should be taken as something separate, defined in terms of 'syllable structure', i.e. in a way prosodically, cf. Williamson 1977. The other useful major class features seem to me still to be "sonorant" and "consonantal". On this point I am unable to follow Williamson, who renounces both of these (1977, 870f), my counterarguments being both that approximants may be voiceless and thus non-sonorants, and that "consonantal" does not concern syllabic function - since e.g. glides are non-consonantal - and should therefore not be integrated into the description of syllable structure.

Lindau 1975 suggests that the frequency of F1 and of F2 - F1 should be used as the features replacing "vowel height" and "backness", respectively. Williamson 1977 argues that "stricture" should distinguish five sound classes: stop, fricative, approximant, high vowel and low vowel, and that sequential articulation should be allowed in the description of e.g. affrication and pre- and post-nasalization (cf. Anderson 1976).

"Consonantal" may be defined as a cover feature (Ladefoged 1971), so that consonantal segments are defined as the complementary class of the intersection of the classes of sonorant, continuant and non-lateral sounds (i.e. [-cons] is equivalent to [+son, +cont, -lat]), cf. Basbøll 1977. Such cover features are used more extensively by Lass (1976) under the name of 'secondary features' which are language specific (whereas the primary features are supposed to be universal). The purpose of these secondary features

is to define 'natural classes' which are useful in the description of a good many phonological processes in one or in several related languages.

Whereas cover features may be seen as abbreviations for sets of features (also cf. Anderson 1974 on glottal features), a possible ordering of the set of features has been discussed too, mainly in terms of hierarchies of strength (recently, e.g., by Hooper 1976 and Foley 1977). One main motivation for proposing these strength hierarchies, which are rooted in the sonority structure of the syllable (ultimately in degrees of physiological opening), is to account for phonotactics (cf. Basbøll 1977), but a lot of evidence from different modalities has been brought into the discussion (for a good critical overview, see Drachman 1977). Broecke has treated hierarchies and rank orders in distinctive features in a monograph (1976).

In addition to the simultaneous (or even paradigmatic) organization of features just mentioned, there exists of course the important temporal organization usually referred to as the syllable. Problems of the phonological syllable have been alluded to above (e.g. in the present and the preceding section), but a few articles on this topic could be mentioned here: The papers e.g. by Bell, Hooper and Vennemann presented at the symposium on the syllable in Boulder, Colorado, in October 1976 (not yet published, as far as I know), the work of Perry and of Kahn, and the discussion of syllabification in French as presented e.g. in Rudes 1976, Selkirk 1978 (who builds upon Liberman and Prince 1977, cf. note 15), Cornulier 1978, and Basbøll forthcoming.<sup>18</sup>

### 3.3 Marking

Although the Prague school notion of markedness in phonology has not been within the central field of investigation since the last congress, neither within the generative school (cf. the revival of the concept by Postal 1968 and SPE), nor outside, it has nevertheless been discussed and used in an interesting way by a number of scholars.

18) Since there is still a persistent and widespread misuse of syllable boundaries in the literature, even by otherwise careful authors, I should like to emphasize that e.g. rules which nasalize a vowel before a tautosyllabic nasal should be stated with the syllable as their domain, and not with a syllable boundary as their utmost limit to the right, since the latter formulation makes the wrong prediction that a consonant occurring between the nasal and the syllable boundary would block the rule.

An excellent account of the notion is found in Hyman's commendable textbook (1975), and a discussion of the markedness model of standard generative phonology is given by Eliasson (1977), who emphasizes the distinction between the formal approach to markedness used in SPE, and an external (or "substance based", cf. section 1.2) approach.

To Bailey (e.g. 1977a), markedness is a crucial concept. He discusses the two 'Greenbergian' (and 'Jakobsonian', one could add) principles: '(i) what is more marked changes to what is less marked', (ii) 'what is less marked is implied by (the presence of) what is more marked' in connection with a lot of data from speech variation (in the broad sense), including both "natural" changes and "unnatural" ones (which are very frequent, e.g. due to borrowing). In his account he makes use of the notion of 'feature weighting', i.e. the features do not form an unordered set, but may be weighted in different ways for different groups of languages (in different periods), e.g. "continuant" is a "heavier" feature with respect to "voice" in Romance (p > b > v) than in Germanic (p > f > v). On phonological "chains" and their relation to markedness, also cf. Fox 1976.

The notion of feature weighting (except for the terminology) has also been used by Henning Andersen (whose work belongs equally to the preceding and the present section) in connection with markedness in vowel systems (1975), and for another typological purpose in (forthcoming), viz. to distinguish between "vocalic" and "consonantal" languages (with different weighting of these features) while exploring a number of consequences (from sound change, etc.) of this typological distinction.

The concepts of markedness, neutralization and archiphonemes are, historically at least, very much connected, cf. the next section.

### 3.4 Archisegments

In the natural generative phonology of e.g. Hooper (1975, 1976) and Rudes (1976), the lexical entries<sup>19</sup> consist of incompletely specified segments ("archisegments") such that all redundant features, both those that represent neutralized contrasts and those that are never contrastive in segments of a given type, are left blank in the lexical representations.

19) The lexical entries consist of whole words according to Rudes, whereas Hooper takes productive suffixes to be separate entries.

The term "archisegment" is formed on analogy with the Praguean "archiphoneme", and it is not surprising that the discussion of the notions of archiphoneme, neutralization and defective distribution has been most lively in a Prague-like functional tradition. Vion 1974 distinguishes between different degrees of relevance for a neutralizable opposition, and Akumatsu 1975 rejects Trubetzkoy's rather abstract notion of a "representative" of an archiphoneme.

Davidson-Nielsen in his monograph (1978), basing his claims upon e.g. speech error evidence and orthographic evidence, defines neutralization as contextually determined (in a purely phonetic/phonological sense) loss of one distinctive dimension (with some further qualifications). By an archiphoneme he understands a contrastive segment in weak position whose distinctive features correspond to the intersection of two contrastive segments in strong position which differ in terms of one feature only.

#### 4. Concluding remarks

As mentioned at the beginning of this report, I am fully aware of the subjectivity of what I have written, both as regards selection and evaluation.<sup>20</sup> Many works of a general nature which are also relevant for phonology have been ignored (but cf. Tench 1976 for an interesting tagmemic account), and many problems and trends have not been considered.<sup>21</sup> Although I have in many places expressed my scepticism about overly abstract approaches to phonology, I should like to state that linguistic generalizations presuppose abstractions, and that extremely concrete phonetic experiments alone do not lead to an adequate understanding of phonological issues. The field of theoretical phonology has not been reduced to any type of orthodoxy. It is still very much alive.

20) It is evident that the task is an infinite one, but I should nevertheless like to emphasize that I know many of the references only superficially.

21) An important problem which has not been discussed is how to settle an underlying form within generative phonology, cf. Zwicky 1975. An example of a trend which has not been covered here is the "atomic phonology", see e.g. Dinnsen and Eckman 1978. A combined example is Hervey 1978 on accidental vs. structural gaps within a functionalist framework.

#### References

- Akamatsu, T. (1975): "De la notion de "représentant de l'archiphonème", in Actes du deuxième colloque de linguistique fonctionnelle, 93-101, Clermont-Ferrand.
- Andersen, H. (1975): "Markedness in vowel systems", Proc.Ling. 11, vol. II, 891-897.
- Andersen, H. (forthcoming): "Vocalic and consonantal languages", in Studia Linguistica A.V. Issatchenko a collegis et amicis oblata, L. Durevic and D.S. Worth (eds.), Lisse: P. de Ridder Press.
- Anderson, J.M. and C. Jones (1974): "Three theses concerning phonological representations", JL 10, 1-26.
- Anderson, S.R. (1974): The Organization of Phonology, New York: Academic Press.
- Anderson, S.R. (1975): "On the interaction of phonological rules of various types", JL 11, 39-62.
- Anderson, S.R. (1976): "Nasal consonants and the internal structure of segments", Lg. 52, 326-344.
- Bach, E. (1968): "Two proposals concerning the simplicity metric in phonology", Glossa 2, 128-149.
- Bailey, C.-J.N. (1976): "Phonology since generative phonology", Papiere zur Linguistik 11, 5-19.
- Bailey, C.-J.N. (1977a): "Linguistic change, naturalness, mixture, and structural principles", Papiere zur Linguistik 16, 6-73.
- Bailey, C.-J.N. (1977b): System of English Intonation with Gradient Models, Bloomington, Indiana: Indiana University Linguistics Club.
- Basbøll, H. (1977): "The structure of the syllable and a proposed hierarchy of distinctive features", in Dressler and Pfeiffer 1977, 143-148.
- Basbøll, H. (1978): "On the use of "Domains" in phonology", Proc. Ling. 12.
- Basbøll, H. (forthcoming): "Schwa, jonctures et syllabification dans les représentations phonologiques du français", ALH 16.2.
- Bolozky, S. (1977): "Fast speech as a function of tempo in natural generative phonology", JL 13, 217-238.
- Brasington, R.W.P. (1976): "On the functional diversity of phonological rules", JL 12, 125-152.
- Brink, L. and J. Lund (1975): Dansk Rigsmål 1-2, Copenhagen: Gyldendal.
- Broecke, M.v.d. (1976): Hierarchies and rank orders in distinctive features, Assen/Amsterdam: Van Gorcum.
- Chen, M. and W. Wang (1975): "Sound change: actuation and implementation", Lg. 51, 255-281.
- Chomsky, N. and M. Halle (1968): The Sound Pattern of English, New York: Harper and Row.

- Clements, G.N. (1977): "The autosegmental treatment of vowel harmony", in Dressler and Pfeiffer 1977, 111-119.
- Cornulier, B. (1978): "Syllabe et suite de phonèmes en phonologie du français", in *Etudes de phonologie française*, B. Cornulier and F. Dell (eds.), 31-69, Paris: Editions du CNRS.
- Davidsen-Nielsen, N. (1978): *Neutralization and Archiphonemes: Two Phonological Concepts and their History*, Copenhagen: Akademisk Forlag.
- Derwing, B. (1973): *Transformational grammar as a theory of language acquisition*, Cambridge Studies in Linguistics 10, Cambridge University Press.
- Dinnsen, D.A. and F.R. Eckman (1978): "Some substantive universals in atomic phonology", *Lingua* 45, 1-14.
- Donegan, P.J. and D. Stampe (forthcoming): "The study of natural phonology", in *Current Approaches to Phonological Theory*, D. Dinnsen (ed.), Bloomington, Indiana: Indiana University Press.
- Drachman, G. (1977): "On the notion 'phonological hierarchy'", in Dressler and Pfeiffer 1977, 85-102.
- Dressler, W.U. (1977a): "Morphologization of phonological processes", in *Linguistic Studies offered to Joseph Greenberg*, A. Juilland (ed.), Saratoga, Calif.: Anna Libri and Co.
- Dressler, W.U. (1977b): Review of Skousen 1975, *Lingua* 42, 223-225.
- Dressler, W.U. and O.E. Pfeiffer (1977), eds.: *Phonologica 1976 = Innsbrucker Beiträge zur Sprachwissenschaft 19*, Innsbruck.
- Eliasson, S. (1977): "Notes on the markedness model of standard generative phonology", in *Provincial and Universal Linguistic Themes*, R. Otterbjörk and S. Sjöström (eds.), 196-219, Publications from the Dept. of Gen. Ling. 12, Univ. of Umeå.
- Eliasson, S. (forthcoming): "Analytic vs. synthetic aspects of phonological structure", in *Phonology in the 1970's*, D.L. Goyvaerts (ed.), Ghent: Story-Scientia.
- Fischer-Jørgensen, E. (1975a): *Trends in Phonological Theory. A Historical Introduction*, Copenhagen: Akademisk Forlag.
- Fischer-Jørgensen, E. (1975b): "Perspectives in Phonology", *Annual Report of the Institute of Phonetics, University of Copenhagen* 9, 215-236.
- Foley, J. (1977): *Foundations of theoretical Phonology*, Cambridge Studies in Linguistics 20, Cambridge University Press.
- Fox, A. (1976): "Problems with phonological chains", *JL* 12, 289-310.
- Goldsmith, J. (1976): "An overview of autosegmental phonology", *Linguistic Analysis*, 23-68.
- Griffen, T.D. (1976): "Toward a nonsegmental phonology", *Lingua* 40, 1-20.
- Griffen, T.D. (1977): "German [x]", *Lingua* 43, 375-390.
- Gussmann, E. (1976): "Recoverable derivations and phonological change", *Lingua* 40, 281-303.
- Herbert, R.K. (1977): "Phonetic analysis in phonological description: Prenasalized consonants and Meinhof's Rule", *Lingua* 43, 339-373.
- Hervey, S.G.J. (1978): "On the extrapolation of phonological forms", *Lingua* 45, 37-64.
- Hockett, C.F. (1975): *A Manual of Phonology*, Indiana University Publications in Anthropology and Linguistics, Memoir II, Baltimore.
- Hooper, J. (1975): "The archisegment in natural generative phonology", *Lg.* 51, 536-560.
- Hooper, J. (1976): *An Introduction to Natural Generative Phonology*, New York etc.: Academic Press.
- Hurford, J.R. (1977): "The significance of linguistic generalizations", *Lg.* 53, 574-620.
- Hyman, L.M. (1975): *Phonology. Theory and Analysis*, New York etc.: Holt, Rinehart and Winston.
- Jakobson, R., M. Halle and G. Fant (1952): *Preliminaries to Speech Analysis*, Cambridge, Mass.: MIT-Press.
- Kiparsky, P. (1973): "Abstractness, opacity and global rules", in *Three Dimensions of Linguistic Theory*, O. Fujimura (ed.), 57-86, Tokyo: TEC Company Ltd.
- Koster, J., H. v. Riemsdijk and J.R. Vergnaud (1978): "GLOW manifesto", *Glow Newsletter* 1, 2-5 (Linguistics Dept., Univ. of Amsterdam).
- Koutsoudas, A. (1977): "On the necessity of the morphophonemic-allophonic distinction", in Dressler and Pfeiffer 1977, 121-126.
- Ladefoged, P. (1971): *Preliminaries to Linguistic Phonetics*, Chicago and London: The University of Chicago Press.
- Ladefoged, P. (1975): *A Course in Phonetics*, New York: Harcourt Brace Jovanovich.
- Lass, R. (1976): *English Phonology and Phonological Theory*, Cambridge Studies in Linguistics 17, Cambridge University Press.
- Leben, W. (1976): "The Tones in English Intonation", *Linguistic Analysis* 2, 69-107.
- Leben, W. and O. Robinson (1976): "'Upside-down' phonology", *Lg.* 53, 1-20.
- Lieberman, M. (1975): *The Intonational System of English*, Bloomington, Indiana: Indiana University Linguistics Club.
- Lieberman, M. and A. Prince (1977): "On Stress and Linguistic Rhythm", *Linguistic Inquiry* 8, 249-336.
- Lindau, M. (1975): *Features for Vowels*, UCLA Working Papers in Phonetics 30, Los Angeles.



- Linell, P. (1977): "Morphophonology as part of morphology", in Dressler and Pfeiffer 1977, 9-20.
- Linell, P. (forthcoming): Psychological Reality in Phonology, Cambridge University Press.
- Mermelstein, P. (1973): "Articulatory model for the study of speech perception", JASA 53, 1070-1083.
- Milner, J.-C. (1978): Reply to Koster et al. 1978, Glow Newsletter 1, 5-9 (Linguistics Dept., Univ. of Amsterdam).
- Pellowe, J. (1976): "The Tyneside linguistic survey: aspects of a developing methodology", in Sprachliches Handeln - Soziales Verhalten: ein Reader zur Pragmalinguistik und Soziolinguistik, W. Viereck (ed.), 203-217 and 365-367, Munich: Wilhelm Fink.
- Pellowe, J. and V. Jones (1978): "On intonational variability in Tyneside speech", in Sociolinguistic Patterns in British English, P. Trudgill (ed.), 101-121, London: Edward Arnold.
- Postal, P. (1968): Aspects of Phonological Theory, New York etc.: Harper and Row.
- Ringen, C. (1976): "Vacuous application, iterative application, re-application and the unordered rule hypothesis", in The Application and Ordering of Grammatical Rules, A. Koutsoudas (ed.), 55-75, The Hague: Mouton.
- Rischel, J. (1964): "Stress, juncture, and syllabification in phonemic description", Proc.Ling. 9, 85-93.
- Rischel, J. (1972): "Compound stress in Danish without a cycle", Annual Report of the Institute of Phonetics, University of Copenhagen 6, 211-230.
- Rischel, J. (1974): Topics in West Greenlandic Phonology, Copenhagen: Akademisk Forlag.
- Rischel, J. (1978): "Some remarks on realism in current phonological work", in Papers from the Fourth Scandinavian Conference of Linguistics, K. Gregersen et al. (eds.) 419-431, Odense University Press.
- Rudes, B.A. (1976): "Lexical representation and variable rules in natural generative phonology", Glossa 10, 111-150.
- Sanders, G. (1977): "On the notions 'optional' and 'obligatory' in linguistics", Ling. 195, 5-47.
- Schane, S. (1968): French Phonology and Morphology, Cambridge, Mass.: MIT Press.
- Selkirk, E.O. (1978): "Comments on Morin's paper: The French foot: on the status of 'mute' e", Studies in French Linguistics 1.2, 141-150 (Bloomington, Indiana: Indiana University Linguistics Club).
- Skousen, R. (1975): Substantive Evidence in Phonology. The Evidence from Finnish and French = Janua Linguarum, series minor, 217, The Hague: Mouton.
- Spang-Hanssen, H. (1959): Probability and Structural Classification in Language Description, Copenhagen: Rosenkilde og Bagger.
- SPE = Chomsky and Halle (1968).
- Stampe, D. (1973): "On Chapter Nine", in Issues in Phonological Theory, M.I. Kenstowicz and C. Kisseberth (eds.), 44-52, The Hague: Mouton.
- Stanley, R. (1967): "Redundancy rules in phonology", Lg. 43, 393-436.
- Tench, P. (1976): "Double ranks in a phonological hierarchy?", JL 12, 1-20.
- Vennemann, T. (1974): "Words and syllables in natural generative grammar", Natural Phonology Parasession, 346-374, Chicago Linguistic Society.
- Vion, R. (1974): "Les notions de neutralisation et d'archi-phonème en phonologie", La linguistique 10.1, 33-52.
- Williamson, K. (1977): "Multivalued features for consonants", Lg. 53, 843-871.
- Zwicky, A.M. (1975): "Settling on an underlying form: The English inflectional endings", in Testing Linguistic Hypotheses, Cohen and Wirth (eds.), 129-185, New York: Wiley.



## NOTES ON THE DEVELOPMENT OF PHONOLOGICAL THEORY

Stephen R. Anderson, Dept. of Linguistics, U.C.L.A., Los Angeles, California, U.S.A.

In describing the state of phonological theory in recent years, Basbøll distinguishes between "substance based" and "formal" approaches to the fundamental problems of definition and explanation in the field. This is undoubtedly a useful opposition, and one that corresponds to most people's intuition about what is at issue in some recent controversies. My own work has been primarily in the direction Basbøll would characterize as "formal", and I would like therefore to describe the issues involved from that perspective. I am sure my fellow co-reporter will do justice to the other side.

I would agree with Basbøll that most phonologizing in recent years has been carried out within a comparatively unitary set of assumptions about the defining problems of the field, and therefore that a single broadly construed school of phonology has dominated research (despite efforts to promote comparatively minor differences of opinion to the status of fundamental differences). Whether acknowledged or not, most of the problems dealt with in this school are set (or at least foreshadowed) in the 'standard theory' of Chomsky and Halle's Sound Pattern of English (SPE). I have given elsewhere an account of recent developments in connection with the details of that program (cf. Anderson, 1979), and will not repeat that discussion here. I will instead confine myself to some remarks of a more general nature.

To gain perspective on the issues involved in recent phonological debate, it seems to me quite useful to consider the parallels between the evolution of phonological theory and that of the study of the foundations of mathematics. Let us recall that the primary nature of a phonological theory, as expressed in SPE, is the development of an explicit formal notation for phonological description. In combination with an evaluation function for grammars defined over this notation, this would result in a comprehensive axiomatization of the subject matter of phonology, in the sense that all problems connected with the discovery of a correct (or 'descriptively adequate') account of sound structure in a given language would thereby be reduced to the mechanical manipulation of expressions in a fully explicit notational system. Of

course, SPE does not claim to have accomplished this goal, but it is nonetheless the program of the theory. The successes achieved within this framework were seen as confirmation of the plausibility of such an axiomatization.

The program of SPE is thus strikingly similar to that of another fundamental work of 20th century thought, Whitehead and Russell's Principia Mathematica (PM). That work developed a program of reducing all of the intellectual content of mathematics to the formal manipulation of expressions in a logistic system by means of fully explicit rules. While the calculus of formal logic in which PM proposed to express mathematical propositions is of course quite unlike the descriptive apparatus for phonological expressions envisaged by SPE, the goal of expressing all of the content of a field in terms subject to formal manipulation by well-established rules is common to the two works.

PM's account of the foundations of mathematics was initially greeted enthusiastically, since it promised to give a full reconstruction of the traditional notion that the truth of mathematical propositions derives from logic alone, and not from contingent facts about the world. This enthusiasm rapidly gave way to dissatisfaction, however, as it became apparent that there were fundamental obstacles to the logicist program. In particular, the theory in its basic form was seen to give rise to a number of the paradoxes which had long troubled mathematicians (such as various forms of the problem of the barber who shaves everyone who does not shave himself, and others). In order to remedy this difficulty, Russell had proposed what is known as the theory of 'types', roughly speaking a restriction on the kinds of classes that can be referred to in a given expression. Unfortunately, the theory of types itself had the undesirable consequence of rendering unstatable or meaningless many basic propositions in number theory. It was thus necessary, in the full system of the PM, to appeal to an axiom of infinity and an axiom of reducibility, whose plausibility and intuitive appeal are vastly less than that of the rest of the logical system. Since the theory of types seemed unavoidable in the context of the logic of the PM, and since it seemed to lead to such counterintuitive emendations of the system, the logicist program for the foundations of mathematics was gradually abandoned.

Partially in response to the perceived failure of this approach, other views of the foundations of mathematics were developed on other assumptions. Among the most important of these alternative views was that presented by Brouwer and others under the title of intuitionism. A primary tenet of this school is the rejection of all expressions purporting to refer to objects that cannot in fact be fully constructed. In particular, expressions that refer to explicitly infinite sets are disallowed, since (while one can give directions for indefinitely enlarging the extension of a set) it is obviously not possible to complete the enumeration of such an object. This has the immediate consequence that the fundamental paradoxes that arise for Russell's system are avoided, since the problematic classes turn out to be impossible to construct within the limits of an intuitionist logic.

Intuitionists have attempted to reconstruct as much as possible of the subject matter of mathematics while adhering to such limitations. In many cases, it turns out to be possible to reformulate classical results in such a way as to be able to derive them in these terms. In other areas, however, this is impossible, and the intuitionists are then led to conclude that such areas of mathematics are in fact meaningless: a somewhat controversial result.

In the course of developing the intuitionist program, its practitioners have clearly revealed much about the conceptual basis of mathematical propositions. This program does not really lead to independent advances, however, since it provides the basis for only a partial development of mathematics. Relatively few working mathematicians seem willing to accept the limitations on their subject matter imposed by the premises of intuitionist logic, and thus although they can be said to have shed light on a (proper) subset of the field, the intuitionists cannot be said to have replaced the traditional modes of inference for mathematics as a whole.

A similar development can be traced in phonology. In particular the program of SPE leads, in the end, to the result that considerations of the substantive phonetic content of representations and rules has no natural role in the system of phonology. This problem is recognized in the famous chapter 9 of SPE, where a solution is proposed in the form of the theory of markedness.

Such a theory is in fact an attempt to reduce exhaustively the considerations of phonetic content that might be relevant to phonology to purely formal expression in the notation. While it too was greeted with much initial enthusiasm, it is noteworthy that essentially no substantial analyses of phonological phenomena have appeared subsequently in which this aspect of the theory plays a fundamental role. This seems to be due at least in part to the fact that the set of 'marking conventions' required to account for the facts of one language or group of languages simply do not extend to comparable utility in others. The purely mechanical problems encountered here are immediately apparent to anyone attempting to formulate a description in such a way, and as a result serious efforts to take account of phonetic content have generally been pursued along quite different lines.

If we would draw the full lesson from these observations, it seems to me that we must conclude that the role of phonetic content in phonology is such as to reveal a fundamental inadequacy in the full 'logician' program for the field sketched in SPE. The theory of markedness, that is, seems to be an emendation of the same character as Russell's theory of types. The lesson in each case is not that a consistent formal system of the required character cannot be constructed, but rather that the only available ways of doing so inevitably lead to fundamental conflicts with the subject matter which the theories are intended to account for. Neither a logical basis for mathematics nor a comprehensive notation for the expression and comparison of phonological descriptions are proven to be wrong: they are simply shown to be incomplete in essential aspects as full reconstructions of the domains of thought with which they are concerned.

In reaction to the inadequacies of the account of phonetic substance offered by SPE, a similar 'intuitionist' approach (though not really in the form of a coherent school) has grown up in phonology, in attempts to remedy the presumed paradoxes resulting from the standard theory by restricting its conceptual richness. Most notably, the approach of Natural Generative Phonology (NGP) has been to require the reconstruction of phonological accounts without appeal to abstract entities or to putatively counterintuitive logistic principles such as relevant explicit ordering. This constitutes a retreat from idealism to a theory

founded insofar as possible on what are (from a linguist's point of view, if not that of an experimental psychologist) the observable and immediately verifiable aspects of linguistic structure. As such, it is immediately reminiscent of the constructivist basis of intuitionist mathematics.

In fact, the parallel is quite close. NGP succeeds in reconstructing a large part of the traditional domain of phonological description, though sometimes in unfamiliar terms. In doing so, it has shown us much about the conceptual basis of more familiar solutions. On the other hand, there are also many aspects of what has usually been taken to be phonology which are inaccessible on its premises. These areas of phonology are either written off altogether (that is, declared to be linguistically meaningless) or ascribed to the operation of vague, nonphonological principles (such as 'via-rules', essentially a name for the description of those aspects of phonology that cannot be accounted for without an appeal to abstract entities).

Now a consistent adherent of NGP may well be happy with the result that certain domains are thereby eliminated from consideration, just as a confirmed intuitionist may be convinced of the result that much of classical and modern mathematics is literally meaningless, but in both areas traditional, pre-systematic practitioners of these subjects have felt discontent with the portion of their fields that can be treated within such a radically 'constructivist' account. If NGP must, as argued in critiques such as that of Gussmann (1978), throw out the baby with the bath water, many phonologists would resist the contention that a priori considerations of psychological reality make this way of avoiding the disregard of phonetic substance characteristic of SPE the right line.

Now in mathematics, the disillusionment with the full logicist program which followed from certain aspects of the system of PM certainly did not have the result that serious work in formal mathematical logic came to a halt. On the contrary, the sort of investigation carried out in these terms turned out to constitute an interesting and coherent field of study, defining significant problems of its own to which solutions could be sought that would result in essential contributions to our understanding of the structure of mathematics. If it is not possible to decide all

mathematical questions within this field, it is still an area of basic importance, concerned with very real problems.

It seems to me that the situation in phonology is entirely analogous. The formalist program of SPE is undoubtedly incomplete as the basis of a comprehensive account of all problems in phonological structure in natural language. It still appears to constitute a well-formed and important subpart of that study, with real problems in its own right that can be formulated, addressed, and decided, and which lead to basic improvements in our understanding of the nature of sound systems. It is in this area, indeed, that I think we are still (largely due to the monumental results represented by SPE) best equipped to make substantial progress. Our growing awareness of the range of problems that cannot be reduced to notational decisions, in fact, has the effect of refining our understanding of the contribution made by those results that can be obtained. In this respect, my own (admittedly quite partisan) evaluation is that the advances that can be made by taking formal questions seriously far exceeds the interesting but limited scope of reductionist efforts such as that of NGP.

As an example of such a question, let us briefly consider the problem of whether or not morpholexical ('word formation') processes necessarily precede purely phonological processes in grammars. It should be emphasized that the notion "precedes" in this formulation of the issue is not a purely metaphoric (or metaphysical) one, nor is its validity dependent on an interpretation in terms of temporal sequential processing, either in speakers' production or in historical change. Rather, it refers to the issue of whether or not there are ever morpholexical processes whose operation crucially depends on (and thus presupposes the presence of) information about a form which is only supplied by the generalizations represented by some phonological process - and which is thus unavailable in the underlying representations of forms. The device of sequential application is a particular formalization of this, but it should be kept in mind that it is the relation of informational dependency that is at issue.

The value of this observation for our knowledge of language, however, turns on the fact that it is logically a contingent proposition. Simply asserted by fiat, it becomes totally uninteresting, a limitation on what sort of world we are willing to countenance.

Taken otherwise, however, it can be falsified by the demonstration that in at least one language there is a well-supported instance of a contrary dependency. Such examples are not, in fact, especially difficult to document. A particularly interesting (because highly structured) case is found in Javanese (cf. Dudas, 1974). In this language, the 'elative' (a sort of intensive form) of adjectives is constructed by replacing the last vowel of the word by a tense high vowel: i if the basic vowel was front and non-round, u if the basic vowel was back and round. Thus we find alternations such as luwe 'hungry', elative luwi; adoh 'far', elative aduh, and many others. If the final vowel is a, however, there are two cases: if the last syllable is closed, the elative is formed in i, as in gampang 'easy', elative gamping. If the last syllable is open, however, the elative is formed in u. Thus, from underlying /kamba/ 'insipid', the elative is kambu. The explanation of this difference is not far to seek, however. A general phonological rule of the language neutralizes the opposition between /a/ and /o/ in final open syllables, replacing both by o. This rule is responsible for alternations such as dino 'day', dinane 'the day' (from the root /dina/), and is dependent only on phonological information for its operation. There is much more to be said about these rules, and about others with which they interact, but I think those who consult Dudas' paper and the sources to which she refers will find that this account does not distort the situation. Now in fact the behavior of basic /a/ in elative formations is clear: it is precisely where this vowel would be replaced by o (in final open syllables) that elative formation treats it in the same way as back rounded vowels (like /o/). Otherwise, it behaves like the other unrounded vowels. The generalization that is apparent in these data is that elative formation depends on the information that is supplied by the rule replacing final /a/ in open syllables by o, not on the underlying form directly. In other words, this rule of word-formation follows the phonological rule in question (as well as some others, as Dudas documents). Notice that this demonstration proceeds quite otherwise than by "considering the ... notation as given ... [and] drawing conclusions ... from the notation" as Basbøll seems to suggest. Rather, it is precisely the appropriate form of the notation that is at issue: in particular, an aspect of the organization of grammars concerning

mathematical questions within this field, it is still an area of basic importance, concerned with very real problems.

It seems to me that the situation in phonology is entirely analogous. The formalist program of SPE is undoubtedly incomplete as the basis of a comprehensive account of all problems in phonological structure in natural language. It still appears to constitute a well-formed and important subpart of that study, with real problems in its own right that can be formulated, addressed, and decided, and which lead to basic improvements in our understanding of the nature of sound systems. It is in this area, indeed, that I think we are still (largely due to the monumental results represented by SPE) best equipped to make substantial progress. Our growing awareness of the range of problems that cannot be reduced to notational decisions, in fact, has the effect of refining our understanding of the contribution made by those results that can be obtained. In this respect, my own (admittedly quite partisan) evaluation is that the advances that can be made by taking formal questions seriously far exceeds the interesting but limited scope of reductionist efforts such as that of NGP.

As an example of such a question, let us briefly consider the problem of whether or not morpholexical ('word formation') processes necessarily precede purely phonological processes in grammars. It should be emphasized that the notion "precedes" in this formulation of the issue is not a purely metaphoric (or metaphysical) one, nor is its validity dependent on an interpretation in terms of temporal sequential processing, either in speakers' production or in historical change. Rather, it refers to the issue of whether or not there are ever morpholexical processes whose operation crucially depends on (and thus presupposes the presence of) information about a form which is only supplied by the generalizations represented by some phonological process - and which is thus unavailable in the underlying representations of forms. The device of sequential application is a particular formalization of this, but it should be kept in mind that it is the relation of informational dependency that is at issue.

The value of this observation for our knowledge of language, however, turns on the fact that it is logically a contingent proposition. Simply asserted by fiat, it becomes totally uninteresting, a limitation on what sort of world we are willing to countenance.

Taken otherwise, however, it can be falsified by the demonstration that in at least one language there is a well-supported instance of a contrary dependency. Such examples are not, in fact, especially difficult to document. A particularly interesting (because highly structured) case is found in Javanese (cf. Dudas, 1974). In this language, the 'elative' (a sort of intensive form) of adjectives is constructed by replacing the last vowel of the word by a tense high vowel: i if the basic vowel was front and non-round, u if the basic vowel was back and round. Thus we find alternations such as luwe 'hungry', elative luwi; adoh 'far', elative aduh, and many others. If the final vowel is a, however, there are two cases: if the last syllable is closed, the elative is formed in i, as in gampang 'easy', elative gamping. If the last syllable is open, however, the elative is formed in u. Thus, from underlying /kamba/ 'insipid', the elative is kambu. The explanation of this difference is not far to seek, however. A general phonological rule of the language neutralizes the opposition between /a/ and /o/ in final open syllables, replacing both by o. This rule is responsible for alternations such as dino 'day', dinane 'the day' (from the root /dina/), and is dependent only on phonological information for its operation. There is much more to be said about these rules, and about others with which they interact, but I think those who consult Dudas' paper and the sources to which she refers will find that this account does not distort the situation. Now in fact the behavior of basic /a/ in elative formations is clear: it is precisely where this vowel would be replaced by o (in final open syllables) that elative formation treats it in the same way as back rounded vowels (like /o/). Otherwise, it behaves like the other unrounded vowels. The generalization that is apparent in these data is that elative formation depends on the information that is supplied by the rule replacing final /a/ in open syllables by o, not on the underlying form directly. In other words, this rule of word-formation follows the phonological rule in question (as well as some others, as Dudas documents). Notice that this demonstration proceeds quite otherwise than by "considering the ... notation as given ... [and] drawing conclusions ... from the notation" as Basbøll seems to suggest. Rather, it is precisely the appropriate form of the notation that is at issue: in particular, an aspect of the organization of grammars concerning

the formalization of possible interdependences between rules. In this case, the answer seems clear. The proposed constraint is not a valid one, and must be replaced by some other, less restrictive (and hence, less interesting) one. The relation between such questions of formalism and the data of actual languages is quite direct, as is the contribution their resolution can make to our understanding of the organization of sound systems.

In contrast to this situation, however, the problem of how phonetic substance is related to formal description will only receive a serious answer when we recognize the possibility of a radical difference between them. In particular, the requirement that in order to have merit, a theory must be explanatory in the sense of being rigidly predictive imposes in essence the requirement that all questions of substance be expressible ultimately in a formal calculus manipulated by mechanical rules of inference. This sort of program, typified by the theory of markedness, has gotten more and more vague of late, but the requirement of predictability amounts to the demand that substance be reduced to a form commensurate with other, formalizable constituents of a phonological description.

It seems to me that this sense of predictability is inappropriate. The existence of distinct linguistic systems developed from a common antecedent through the differential operation of historical change, taken seriously, provides a falsification of its premises nearly as fundamental as Gödel's demonstration that there are propositions formulable within arithmetic whose truth value cannot be decided in principle within that system. Appeals to social factors and the like are at present mere hand-waving: the conviction that somewhere an explanation exists that will preserve predictability. We must recognize that it is precisely the character of phonetic substance to be both non-random and non-deterministic: a 'logical' formalization of its role in phonology is unavailable in principle. I have suggested elsewhere an alternative sort of goal, the attainment of an ex post facto understanding of phonological processes (or 'exegetic adequacy'), which is (at least, at present) more appropriate for phonology than the program of complete predictability.

When the principles of a theory lead to a domain of conflict, as for instance in the case of the Neogrammarian notions of Laut-

gesetz and Analogie, we certainly do not have predictability - but that does not mean we have not advanced our knowledge. We may well claim to understand the facts to a greater degree than we would in the absence of principles, despite the fact that we cannot claim that the facts could not have been otherwise. An excellent example of this situation is furnished by the current state of research into apparently well-motivated but mutually inconsistent principles that govern rule orderings in phonology.

The atmosphere of 'science', toward which we all aspire, tends to force us into a rather radical mechanism. This is useful when it makes us examine the conceptual bases of our work and to seek the regular connections among phenomena; but it may ultimately become sterile if we insist that only a completely deterministic account is worthy of consideration as 'scientific'. After all, if physics and mathematics can accept fundamental principles of indeterminacy, phonologists should be willing to countenance the uncertain as well.

Basbøll is surely right that an understanding of the role of substance in phonology can only come from an appreciation of the science of that substance, to wit, phonetics. Equally clear, however, not all of the results of phonetic research are equally applicable. It is an axiom of applied mathematics (though not, it sometimes appears, of all phoneticians) that "the purpose of computing is insight, not numbers"; and the most central sort of phonetic research is undoubtedly that which aims at a notion of phonetic motivation and explanation. The work of scholars such as Sweet, Passy, Grammont, and others of an earlier generation has somewhat fallen out of favor as unscientific, largely because of its non-deterministic character (though also on account of the charge of vagueness).

The most promising sort of synthesis seems to me to be found in the work of Baudouin de Courtenay, the 50th anniversary of whose death we mark this year. Baudouin's inspired integration of the explanatory role of traditional phonetics (in accounting for the entrance of low-level processes into the system) with that of the study of the internal structure of grammars (in treating the relations, both evolutionary and synchronic, among the various sorts of rules) deserves serious reconsideration (cf. deChene and Anderson, 1979). Such a synthesis is also one of the merits of Stampe and Donegan's 'Natural Phonology'.



The kind of understanding we can hope to achieve from phonological research, then, is arguably possible only if we abandon the ultimately unreachable goal of complete predictability. Attempts to achieve predictability by imposing arbitrary limits on the form of phonological descriptions, such as by the decision a priori that 'extrinsic' or language particular orderings are one kind of complexity that languages absolutely cannot tolerate, seem unmotivated and misguided. In the absence of an understanding of general cognitive processes underlying language that could explain them, such 'constraints' cannot be taken seriously as the motivation for particular decisions about the appropriateness of descriptions. As Basbøll notes, such extralinguistic explanation seldom plays a real (rather than rhetorical) role in phonological theorizing.

To me, however, this suggests that much of the actual research Basbøll characterizes as 'substance based' is ultimately unproductive, since it is based on the arbitrary imposition of restrictive principles which rule out otherwise well-motivated descriptions. We have no way of knowing a priori what sorts of complexity, abstractness, etc. are tolerated by natural languages, and the only way of discovering this is through the unbiased examination of the facts they present. This is not to deny that such programs can lead to significant insights, as in the case of their emphasis on a distinction between morphological and purely phonological rules, which has evidently led to major improvements in our understanding of sound systems. Nonetheless, far from suggesting that the study of formal problems in phonology, of the sort arising in the framework of SPE, should be abandoned, the lesson of this research seems to be that it is only by taking these formal matters seriously that their ultimate role in a comprehensive view of sound structure can be appreciated.

#### References

- Anderson, Stephen R. (1979): "On the subsequent development of the 'standard theory' in phonology", in Current approaches to phonological theory, D. Dinnsen (ed.), Bloomington, Indiana: Indiana University Press.
- deChene, Brent, and Stephen Anderson (1979): "On compensatory lengthening", Lg. 55.
- Dudas, Karen (1974): "A case of functional phonological opacity: Javanese relative formation", in Studies in the linguistic sciences 4:2, 91-111.
- Gussmann, Edmund (1978): Explorations in abstract phonology, Lublin: Uniwersytet Marii Curie-Skłodowskiej.

#### FORMAL AND SUBSTANTIVE APPROACHES TO PHONOLOGY

Joan B. Hooper, State University of New York at Buffalo, USA

The main report on phonological theory by Hans Basbøll gives a rather thorough treatment of current phonological research. This response will mention a few additional works and issues, but it is intended primarily to supplement Basbøll's report by presenting in somewhat greater depth an examination of the theoretical diversity underlying current phonological research. Our point of departure is the distinction Basbøll discusses between a "substance based" versus a "formal" approach to phonological research. This distinction characterizes quite broadly two major research trends in generative phonology, but leaves out some important differences. In order to highlight several theoretical positions, the "substance" versus "formal" distinction will be divided into two separate distinctions which cross-classify fully. This brief report will discuss the resulting categories and the type of research emanating from each of them. Basbøll noted that his classification of two types of phonology was only rough and ignored some individual differences. Similarly, the distinctions I will make are also rough, and are meant only as a useful organization of a diversity of research perspectives.

#### 1. Two major issues

1.1. The most direct interpretation of the substance-formal distinction divides phonological research into that which investigates formal or structural properties of grammars and that which investigates substantive properties. The former research is concerned with levels of representation, and how they relate to one another, and with the formal properties of rules, and the formal relations among them. Substantive properties can be thought of roughly as content properties -- phonological features are the content of representations, and changes in phonological features in the presence of other related phonological features are the content of rules. For most investigators, the substance of phonology is phonetic (but see section 3.2.).

This aspect of the substance/formal distinction is not so much a theoretical issue as a distinction between two types of interests, which are not mutually exclusive. Most researchers would agree that phonology has both a formal and substantive side, and that the two need to be studied together at least to some extent.



1.2. The second distinction that divides current work in phonology is a distinction in terms of theory, and thus has more serious consequences. Following common practice, this can be labelled the concreteness versus abstractness issue, although it is not abstractness per se that I will focus on here. There are many different degrees of abstractness. We can find a discrete division on this scale, however, if we consider one issue -- the use of data in analysis, in particular the importance of surface facts. In the transformational tradition, one working hypothesis seems to be that if  $x$  and  $y$  share some characteristics, then they must have the same underlying form. This produces an emphasis on the similarities between elements, and leads to a dismissal of their surface differences. Similarly, the goal of uncovering all the "linguistically significant generalizations" the data can yield makes it desirable to ignore counter-indications on the surface. The contrary position is that the rules of the grammar must be fully compatible with the surface data, and exceptions must be taken as giving evidence of rule productivity or the lack of it. Either of these approaches to the evidence can be combined with an interest in the formal or the substantive aspects of phonology.

## 2. The "abstract" positions

2.1. The tradition of the Sound Pattern of English (Chomsky and Halle 1968) combines the abstract approach to data with largely formal interests. Some new issues have arisen in this framework, such as recoverability, a relative of opacity (Leben 1977, Kaye 1978, as well as the references Basbøll cites), and some of the older issues, such as extrinsic rule order and rule types continue to be discussed (see Basbøll's report). It seems unlikely that these issues will ever be resolved, because of the approach to data customary in this framework. Since there is no requirement that a rule correspond in any predetermined way to the surface data, it is impossible to tell if the rules whose relationships are being studied are indeed rules of the grammar. It must be emphasized that the lack of importance of surface data is not an oversight, but rather is a deliberate component of this point of view, as is clear from the following statement by Keyser 1975. He has just argued for internal structural reasons that there is a rule of metathesis in Old English. He then says: "It is a rule whose output never appears unmodified on the surface. This fact

may lead one to suppose that the rule is, therefore, not a possible rule of phonology. However, such a supposition seems to be based upon an excessive reliance on surface data" (pp. 410-411, emphasis mine, JBH).

Of course, Keyser's position is an extreme one. There are many more concrete works in the same general framework (e.g. Kiparsky and O'Neil's 1976 response to Keyser), and many explicit attempts especially by Kiparsky to make the theory more concrete. Despite Kiparsky's various conditions on grammars (e.g. Kiparsky 1976), his work remains in the same framework because he conceives of the grammar as something only indirectly related to surface data. This is evident in Kiparsky 1974 where apparent surface simplifications that must be represented as grammatical complications are lamented, and where the disparity between surface notions of opacity and paradigm uniformity and the formal notion of simplicity are discussed.

There are some works which explicitly disown the SPE model, while maintaining a similar view of surface data, and an interest in the formal aspects of phonology. One example is Leben and Robinson's "Upside-down phonology" which incorporates a very concrete level of lexical representation, while still allowing the formulation of abstract rules, such as the English vowel-shift rules, which are not disconfirmable by surface facts. It is claimed that this framework eliminates ad hoc exception features, but this strikes me as being of dubious value, since this is accomplished by saying nothing about exceptions at all.

2.2. The abstract approach to data can also be combined with an interest in substance, as illustrated in Chapter 9 of SPE. This particular proposal is probably the least satisfying of substantive proposals, because it was appended to a pre-existing formal machinery, and assumes the correctness of certain features and rules. Further, because of the view of surface data mentioned above, the theory does not generate testable hypotheses. Foley's 1977 approach seems closest to the SPE approach in its abstractness, but his proposals are more sophisticated because of a wider data base, a unified theory and the ability to incorporate more than two values for a given feature.

Neither natural phonology (Stampe 1973, Donegan and Stampe 1977) nor polylectal analysis (Bailey 1973 and 1978, and other

articles, too numerous to cite) imposes strict empirical criteria on what may be a rule. Exceptions to rules can arise through extrinsic ordering without affecting the validity of the rule. In natural phonology the reason is that all processes are universal, and occur in all languages unless they are explicitly suppressed. Thus a language does not have to directly evidence a process in order to have it. A polylectal grammar must be quite abstract if a large number of surface variants are to derive from a common underlying form. Both theories claim a close relation between phonetics and phonology, which gives their proposals an empirical aspect, since hypotheses about phonetic motivation can often be tested. (This is in contrast to Foley's theory, in which it is quite explicit that phonology has nothing to do with phonetics.) Moreover, natural phonology makes the important distinction between natural processes and acquired rules, which delimits the input to a theory of rule naturalness. Bailey's work (e.g. Bailey 1978) deals primarily with the very concrete details of phonetic realizations. Both of these approaches differ from the more formal abstract approaches by recognizing variation, and considering many types of independent evidence.

Donegan and Stampe 1977 have entered the race to invent universal principles of extrinsic rule order, with an interesting twist -- a substantive determination of ordering, by which fortition processes apply before lenition processes. The difficulty here is in dividing all processes up into the two types without seriously distorting some of them.

### 3. The "concrete" positions

The theories treated here as concrete have in common the requirement, either implicit or explicit, that the rules of the grammar represent true generalizations about the surface data. The rules are therefore disconfirmable and serve as solid input to theory development in both formal and substantive concerns.

The formal issues do not include rule order, but do include rule type. The distinction between phonetically-conditioned and morphologically-conditioned rules seems firmly established (Andersen 1969, Vennemann 1971 and Hooper 1976). Klausenberger 1978 compares this distinction to Kruszewski's categories of sound alternations, arguing for a third type of rule corresponding to Kruszewski's third category, which are rules with a general morpho-

logical function. There is some debate concerning Vennemann's 1972 via-rules, with Tiersma 1978 arguing for bidirectional rules, and Leben 1977 arguing for a parsing model.

The issues concerning underlying representations involve the unit of representation, morpheme or word, and the presence or absence of redundant feature specifications. These questions are often argued on purely formal grounds, since substantive evidence about underlying forms is difficult to obtain. However, substantive evidence is presented by Vennemann 1978, who argues from historical data that full paradigms must be listed lexically, and Vincent 1978, who finds historical evidence that at least some paradigms must be listed. With regard to redundant feature specifications only Davidsen-Nielsen 1977 has been able to present firm substantive evidence on this issue, and his evidence argues for archi-phonemic representations. Evidence about rules does not bear directly on underlying forms (as Basbøll implies, footnote 12), but rather these issues must be explored separately (see section 3.2.). Thus, for the moment, we must be content with formal arguments concerning underlying representation, such as those found in Hudson 1978, who argues that rules governing automatic alternations only add feature values, never change them, and Skousen 1977, who argues that constraints on underlying forms must be approached from the point of view of language acquisition.

3.2. There are quite a variety of approaches to substance from a concrete perspective, which I take to be a good sign, since this seems to be one of the most fruitful research perspectives. I begin with three proposals presented at the Bloomington Conference on the Differentiation of Phonological Theories.

Dinnsen 1977, Houlihan and Iverson 1977, and Sanders 1977 all propose theories that attempt to define "possible phonological rule". For the most part, they limit the input to their investigations to surface-true phonologically-conditioned rules, but none state that they would impose this limitation on individual grammars. Rather, it seems this limitation is imposed to make their hypotheses testable. Sanders' proposal concerning possible rules is embedded in the larger (formal) framework he calls Equational Grammar. His claim concerning phonological rules is that the directionality of rules is universal, so that if one language has a rule  $A \rightarrow B$ , no language will contain the converse rule  $B \rightarrow A$  in the

same environment. The directionality is functionally determined. It follows from his Simplex Feature notation that allophonic rules will only add features, producing more marked segments. Neutralization rules, on the other hand, produce phonetic structures that are relatively unmarked and communicatively more valuable (than the structures they apply to). Markedness is determined by universal distribution and communicative value "on the basis of physical, social and psychological efficiency" (p. 27). The specific claims are, e.g. that if one language has prothesis (as for example Spanish does), then no language has apheresis. The case is not convincing in view of the large number of converse processes discussed in Andersen 1972, and the small number of examples given by Sanders.

Houlihan and Iverson 1977 make a similar claim; however, their definition of neutralization contains a built-in "blocking" device. They adopt Kiparsky's 1976 definition which says that a rule is neutralizing only if it produces strings or segments that are identical to some strings or segments that are input to the rule. Since the level of input to the rule is an abstract level of the linguist's own devising, potential counter-examples are easily dismissed. Thus it is claimed that English vowel reduction, which produces schwa in unstressed syllables is not a neutralization, since one can analyze English as having no underlying schwa.

These proposals refer to the structure of contrasts in the system to determine what is a possible rule for the language. This is a formal criterion. The substance involved is markedness. The "naturalist" point of view would oppose this "structuralist" point of view and claim that the processes have their own phonetic teleology, and care little about whether they are neutralizing contrasts in a language or not. It should be further noted that these proposals refer only to the structural change of the rule and say nothing of the environment. It seems to me that the environments are just as important and should be subject to cross-linguistic comparison as in Ferguson 1978, and other articles in Greenberg et al. 1978.

Atomic phonology (Dinnsen and Eckman 1977, Dinnsen 1977) incorporates certain testable claims such as, if fricatives devoice word-finally, stops will also devoice word-finally (the latter is the independent or atomic rule, the former its complement). In

Dinnsen 1978 it is argued that these atomic rules are linguistic primes which are not further analyzable nor explicable. Dinnsen argues explicitly against the position that phonological rules are "phonetically explainable" (as claimed in Hooper 1976). His argument is that different languages have different ways of resolving phonological problems. When tautosyllabic consonants differing in voicing arise by morpheme combination in English, they are subject to a progressive devoicing, while in Catalan, they are subject to regressive voicing.

Of course it is true that it is not possible at present to predict which language will have a certain process, especially on the basis of the kinds of information phonological grammars traditionally include. But it is certainly possible that the processes of a language are dependent upon one another, or on typological properties of the language. It is probably no accident that English and Catalan have different processes, since they also have different syllable structure, different stress and different rhythm. What is called for now are typological studies such as Andersen 1978, and studies that combine typological and phonetic substance. Alan Bell and I had this need in mind when we organized the symposium whose proceedings are contained in Bell and Hooper 1978. The emphasis here is on phonetic, psychological and typological facts that may help us understand the diversity of phenomena associated with the combination of segments into larger units.

Finally, a very exciting new perspective is opening up. This is the possibility of approaching traditionally formal or structural problems from a substantive point of view. Hyman 1977 and Hooper 1977 quite independently come to the conclusion that formal distributional criteria cannot always determine the underlying representations of a language. Hyman gives the historical argument that if what is predictable gradually becomes contrastive, there must be a stage in which a feature is both represented lexically, and predicted by rule. Hooper 1977 presents language acquisition data that shows children treating a "predictable" feature (vowel nasality in English) as contrastive. Implicit in these studies is the notion that there may be some concept of "phonetic distance" that partially or fully determines the speaker's analysis into elements represented lexically and elements predictable by rule (cf. Stampe's notion of "minimal structural change"). This opens

up the possibility of substantive phonetic criteria for phonemic analysis. Along similar lines, Comrie 1976 points out that in several cases the development of exceptions to subparts of rules can be correlated with a greater phonetic change produced by that subpart of the rule. If it is possible that even exceptions are not totally arbitrary, it is all the more important to pay attention to them, and to other surface facts of phonology.

#### References

- Andersen, H. (1969): "A study in diachronic morphophonemics: The Ukrainian prefixes", Language 45, 807-830.
- Andersen, H. (1972): "Diphthongization", Language 48, 11-50.
- Andersen, H. (1978): "Vocalic and consonantal languages", in Studia Linguistica, A.V. Issatchenko a collegis et amicis oblata, L.Durevic and D.S. Worth (eds.), Lisse: P. de Ridder Press.
- Bailey, C.J.N. (1973): Variation and linguistic theory, Arlington, Virginia: Center for Applied Linguistics.
- Bailey, C.J.N. (1978): "Gradience in English syllabization and a revised concept of unmarked syllabization", Indiana University Linguistics Club Publication.
- Bell, A. and J.B. Hooper (eds.) (1978): Syllables and segments, Amsterdam: North Holland.
- Chomsky, N. and M. Halle (1968): The sound pattern of English, New York: Harper and Row.
- Comrie, B. (1976): "Morphophonemic exceptions and phonetic distance", to appear in Linguistics, presented at the LSA Summer Meeting, Oswego, N.Y.
- Davidson-Nielsen, N. (1977): "A phonological analysis of English sp, st and sk with special reference to speech error evidence", Journal of the International Phonetic Association 5, 3-25.
- Dinnsen, D. (1977): "Atomic phonology", to appear in Current approaches to phonological theory, D. Dinnsen (ed.), Bloomington: Indiana University Press.
- Dinnsen, D. (1978): "Phonological rules and phonetic explanation", Indiana University Linguistics Club.
- Dinnsen, D. and F. Eckman (1977): "Some substantive universals in atomic phonology", Lingua 45, 1-14.
- Donegan, P. and D. Stampe (1977): "The study of natural phonology", to appear in Current approaches to phonological theory, D. Dinnsen (ed.), Bloomington: Indiana University Press.
- Ferguson, C.A. (1978): "Phonological processes", in Universals of human language, Volume 2, Phonology, J.A. Greenberg, C.A. Ferguson and E.A. Moravcsik (eds.), 403-442, Stanford: Stanford University Press.
- Foley, J. (1977): Foundations of theoretical phonology, Cambridge: Cambridge University Press.
- Greenberg, J.H., C.A. Ferguson and E.A. Moravcsik (1978): Universals of human language, Volume 2, Phonology, Stanford: Stanford University Press.
- Hooper, J.B. (1976): Introduction to natural generative phonology, New York: Academic Press.
- Hooper, J.B. (1977): "Substantive evidence for linearity: vowel length and nasality in English", in Woodford Beach et al. (eds.), Papers from the Thirteenth Regional Meeting, Chicago: Chicago Linguistic Society, 152-164.
- Houlihan, K. and G. Iverson (1977): "Functionally constrained phonology", to appear in Current approaches to phonological theory, D. Dinnsen (ed.), Bloomington: Indiana University Press.
- Hudson, G. (1978): "Automatic alternations in non-transformational phonology", to appear in Language.
- Hyman, L.M. (1977): "Phonologization", in Linguistic Studies Offered to Joseph Greenberg, A. Juillard (ed.), 407-418, Saratoga, California: Anma Libri.
- Kaye, J. (1978): "Recoverability, abstractness and phonotactic constraints", to appear in Phonology in the 1970's, D. Goyvaerts (ed.), Ghent: Story Scientia.
- Keyser, S.J. (1975): "Metathesis and Old English Phonology", Linguistic Inquiry 4, 377-412.
- Kiparsky, P. (1974): "On the evaluation measure", Natural Phonology Parasession, 328-337, Chicago: Chicago Linguistic Society.
- Kiparsky, P. (1976): "Abstractness, opacity and global rules", in The application and ordering of grammatical rules, A. Koutsoudas, 160-186, The Hague: Mouton.
- Kiparsky, P. and W. O'Neil (1976): "The phonology of Old English inflections", Linguistic Inquiry 7, 527-558.
- Klausenberger, J. (1978): "Nikolaj Kruszewski's theory of morphophonology", Historiographia Linguistica 5, 109-120.
- Kruszewski, N. (1881): Über die Lautabwechslung, Kazań.
- Leben, W.R. (1977): "The phonological component as a parsing device", in Current approaches to phonological theory, D. Dinnsen (ed.), Bloomington: Indiana University Press.
- Leben, W.R. and O.W. Robinson (1977): "'Upside-down' phonology", Language 53, 1-20.
- Sanders, G. (1977): "Equational phonology", to appear in Current approaches to phonological theory, D. Dinnsen (ed.), Bloomington: Indiana University Press.
- Skousen, R. (1977): "Analogical sources of abstractness", to appear in Phonology in the 1970's, D.L. Goyvaerts (ed.), Ghent: Story Scientia.
- Stampe, D. (1973): A dissertation on natural phonology, University of Chicago Doctoral Dissertation.
- Tiersma, P. (1978): "Bidirectional leveling as evidence for relational rules", Lingua 45, 65-77.

- Vennemann, T. (1971): "Natural generative phonology", paper read at annual meeting of the LSA, St. Louis, Missouri.
- Vennemann, T. (1972): "Rule inversion", Lingua 29, 209-242.
- Vennemann, T. (1978): "Rule inversion and lexical storage: the case of Sanskrit visarga", in Recent developments in historical phonology, J. Fisiak (ed.), 391-408, The Hague: Mouton.
- Vincent, N. (1978): "Words versus morphemes in morphological change: the case of Italian -iamo", to appear in Historical Morphology, J. Fisiak (ed.), The Hague: Mouton.

## S p e c i a l   L e c t u r e s

GUNNAR FANT:

The relations between area functions and the acoustical signal

155

OSAMU FUJIMURA:

Modern methods of investigation in speech production

161

NIELS A. LASSEN:

The physiology and pathophysiology of language functions as illustrated by measurements of the regional blood flow in the cortex of the brain

167

HISASHI WAKITA:

New methods of analysis in speech acoustics

169

## THE RELATIONS BETWEEN AREA FUNCTIONS AND THE ACOUSTICAL SIGNAL

Gunnar Fant, Dept. of Speech Communication, Royal Institute of Technology (KTH), S-100 44 Stockholm 70, Sweden

Vocal-tract modeling

What progress have we had in vocal-tract modeling and associated acoustic theory of speech production during the last 20 years? My impression is that the large activity emanating from groups engaged in speech production theory and in signal processing has not been paralleled by a corresponding effort at the articulatory phonetics end. Very little original data on area functions have accumulated. The Fant (1960) Russian vowels have almost been overexploited. Our consonant models are still rather primitive and we lack reliable data on details of the vocal tract as well as of essential differences between males and females and of the development of the vocal tract with age.

The slow pace in articulatory studies is of course related to the hesitance in exposing subjects to X-ray radiation. Much hope was directed to the transformational mathematics for deriving area functions from speech-wave data. These techniques have as yet failed to provide us with a new reference material. The so-called inverse transform generates "pseudo-area functions" that can be translated back to high quality synthetic speech but which remain fictional in the sense that they do not necessarily resemble natural area functions. Their validity is restricted to non-nasal, non-constricted articulations and even so, they at the best retain some major aspects of the area function shape rather than its exact dimensions. However, some improvements could be made, even with respect to the possibility to track a side branch of the vocal tract.

Once a vocal-tract model has been set up it can be used, not only for studying articulation-to-speech wave transformations, but also for a reverse mapping of articulations and area functions to fit specific speech-wave data. These analysis-by-synthesis remapping techniques as well as perturbation theory for the study of the consequences of incremental changes in area functions or of the inverse process are useful for gaining insight in the functional aspect of a model. However, without access to fresh articulatory data the investigator easily gets preoccupied with his basic model and the constraints he has chosen.

The slow advance we have had in developing high quality synthesis from articulatory models is in part related to our lack of

reliable physiological data, especially with respect to consonants, in part to the difficulty involved in modeling all relevant factors in the acoustic production process. The most successful attempt to construct a complete system is that of Flanagan et al. (1975) at Bell Laboratories. A variety of studies at KTH in Stockholm and at other places has contributed to our insight in special aspects of the production process such as the influence of cavity-wall impedance, glottal and subglottal impedance, nasal cavity system, source-filter interaction, and formant damping. These will be dealt with in a separate paper.

#### An example

The area functions of male and female articulations of the Swedish vowels [i] and [u] and corresponding computed resonance mode pattern in Fig. 1 may serve to illustrate some findings and problems. The data are derived from tomographic studies in Stockholm many years ago in connection with the study of Fant (1965, 1966) and were published by Fant (1976). It is seen that in spite of the larger average spacing of formants in the female F-pattern related to the shorter overall vocal tract length, the female  $F_1$  and  $F_2$  of [u] and the  $F_3$  of [i] are close to those of the male. This is an average trend earlier reported by Fant (1975a). Differences in perceptually important formants may thus be minimized by compensations in terms of place of articulation and in the extent of the area function narrowing. Such compensations are not possible for all formants and cannot be achieved in more open ar-

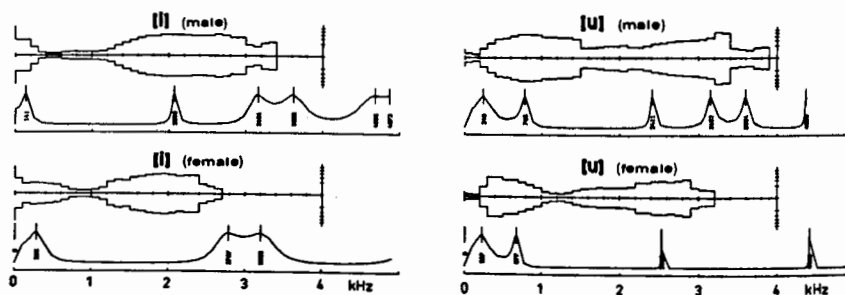


Figure 1

Multicylinder representation of VT area functions of male and female vowels [i] and [u] together with corresponding F-pattern. The shunting effect of sinus piriformis and of cavity walls is not included.

ticulations. The great difference in  $F_2$  of [i] is in part conditioned by the relatively short female pharynx but can in part be ascribed to the retracted place of articulation. It is also disputable whether this particular female articulation serves to ensure an acceptable [i] or whether there is a dialectal trend towards [ɪ]. Also, it is to be noted that X-ray tomography may impede the naturalness of articulations because of the abnormal head position required.

#### Perturbation theory and the inverse transform

Perturbation theory describes how each resonance frequency,  $F_1 F_2 F_3$ , etc. varies with an incremental change of the area function  $A(x)$  at a coordinate  $x$  and allows for a linear summation of shifts from perturbations over the entire area function. The relative frequency shift  $\delta F/F$  caused by a perturbation  $\delta A(x)/A(x)$  is referred to as a "sensitivity function". We may also define a perturbation  $\delta \Delta x/\Delta x$  of the minimal length unit  $\Delta x$  of the area function which will produce local expansions and contractions of the resonator system. It has been shown by Fant (1975b), Fant and Pauli (1974), that the sensitivity function for area perturbations of any  $A(x)$  is equal to the distribution with respect to  $x$  of the difference  $E_{kx} - E_{px}$  between the kinetic energy  $E_{kx} = \frac{1}{2}L(x)U^2(x)$  and the potential energy  $E_{px} = \frac{1}{2}C(x)P^2(x)$  normalized by the totally stored energy in the system.

The distribution of the sum of the kinetic and potential energies describes the sensitivity to length scale perturbations and provides furthermore a realistic quantitative measure of the dependency of the resonance mode on various parts of the area function. Length perturbation has been applied to the problem of scaling the pharynx and the mouth differently, comparing male and female articulations, Fant (1975b).

If the perturbation function is expressed as a function of as many parameters as there are formants, it is possible to calculate the change in area function from one F-pattern to another, Fant and Pauli (1974). This technique has been used by Mrayati and Guérin (1976) for deriving plausible area functions for French vowels on the basis of their deviation from my reference Russian vowels. This procedure must be administered in steps of incremental size with a recalculation of the sensitivity function after each step.



I shall not go into details of the mathematics of the inverse transform. The usual technique, e.g. Wakita (1973), is to start out with a linear prediction (LPC) analysis of the speech wave to derive the reflection coefficients which describe the analog complex resonator. The success of this method is dependent on how well the losses in the vocal tract are taken into account. Till now the assumptions concerning losses have been either incomplete or unrealistic. Also the processing requires that the source function be eliminated in a preprocessing by a suitable deemphasis or by limiting the analysis to the glottal closed period. In spite of these difficulties the area functions derived by Wakita (1973) preserve gross features.

In general, a set of formant frequencies can be produced from an infinite number of different resonators of different length. We know of many compensatory transformations, such as a symmetrical perturbation of the single-tube resonator. However, if we measure the input impedance at the lips, Schroeder (1967), or calculate formant bandwidths, we may avoid the ambiguities. A technique for handling tubes with side branches has been proposed by Ishizaki (1975).

The following very general discussion of the inverse transform is based on a lossy transmission line representation of each section of the area function. The approach is similar to that of Atal et al. (1978).

It can be shown that a number of  $m$  formants, specified by their frequencies and bandwidths potentially define a unique area function with  $2m$  degrees of freedom providing that the resistive elements that determine the bandwidths are unique functions of frequency and of the resonator configuration. It follows that given any total length of an area function, it can be quantized in  $2m$  sections of equal length and there could exist a unique solution for the  $2m$  area values. The non-uniqueness of the overall length may be overcome by adding one more formant to the specification.

Another solution which is unique with respect to vocal-tract length is a configuration of a cascade of  $m$  cylindrical tubes, each specified by area and length derived from the  $m$  formant frequencies and bandwidths. We can exemplify this model by the single-tube resonator. Its length determines the lowest resonance frequency and the area is determined from the bandwidth measure. An F-pattern with  $F_1 = 260$ ,  $F_2 = 1990$ , and  $F_3 = 3050$  Hz appropriate for the

vowel [i] would be generated by a two-tube system in which the back tube has an area of  $8 \text{ cm}^2$  and a length 8.7 cm, and the front tube an area of  $1 \text{ cm}^2$  and effective length 5.8 cm. The compensatory articulation with the same areas but exchange of lengths has exactly the same pattern of all formant frequencies, Fant (1960), but a different bandwidth pattern. A minimum of two frequencies and two bandwidths would theoretically suffice for a unique derivation of either configuration. In practice it may take a ventriloquist to produce both variants. Possibly, the variant with short back cavity would fit the shape of a child's vocal tract. Other aspects of front-back compensations have been treated by Öhman and Zetterlund (1974).

On the whole, we are free to choose any parametric specification to fit a continuous area function providing the number of parameters is twice the number of formants specified in both frequency and bandwidth. Unless the total length is a unique function of the parameters we need one more formant to be specified. We could thus construct a four-tube model with or without smoothing between sections to be uniquely defined by four frequencies and four bandwidths. A combination of this technique with specific constraints, such as a fixed larynx tube, may be introduced to concentrate the predictive capacity to other parts of the system.

This simple reasoning has potentialities to be exploited more than has been done. In practice, however, as pointed out by Atal et al. (1978), we might find that bandwidths may come out the same in two alternative configurations or that their difference may turn out to be smaller than what we can accurately measure. Some additional redundancy could be introduced to overcome such difficulties.

A lack of bandwidth measures can generally not be compensated for by introducing more formant frequencies. On the other hand, if we resort to an articulatory model with natural constraints on possible area functions we may base the prediction on formant frequencies alone, Lindblom and Sundberg (1969), Ladefoged et al. (1978). However, the same pattern of, say,  $F_1$ ,  $F_2$  and  $F_3$  could generate somewhat different area functions in other models with other constraints, e.g. in terms of a different overall length. A combination of formant frequencies, bandwidths and articulatory constraints should be optimal.

References

- Atal, B.S., J.J. Chang, M.V. Mathews, and J.W. Tukey (1978): "Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique", JASA 63, 1535-1555.
- Fant, G. (1960): Acoustic Theory of Speech Production, 's-Gravenhage: Mouton (2nd edition 1970).
- Fant, G. (1965): "Formants and cavities", in Proc. 5th Int. Congr. of Phonetic Sciences, Münster, E. Zwirner (ed.), Basel: Karger.
- Fant, G. (1966): "A note on vocal tract size factors and non-uniform F-pattern scalings", STL-QPSR 4, 22-30 (KTH, Stockholm).
- Fant, G. (1975a): "Non-uniform vowel normalization", STL-QPSR 2-3, 1-19 (KTH, Stockholm).
- Fant, G. (1975b): "Vocal-tract area and length perturbations", STL-QPSR 4, 1-14 (KTH, Stockholm).
- Fant, G. (1976): "Vocal tract energy functions and non-uniform scaling", J.Acoust.Soc.Japan 11, 1-18.
- Fant, G. and S. Pauli (1974): "Spatial characteristics of vocal tract resonance modes", in Speech Communication, Vol. 2, G. Fant (ed.), 121-132, Stockholm: Almqvist & Wiksell 1975 (Proc. SCS-74, Stockholm).
- Flanagan, J.L., K. Ishizaka, and K. Shipley (1975): "Synthesis of speech from a dynamic model of the vocal cords and vocal tract", Bell System Techn. J. 54, 485-506.
- Ishizaki, S. (1975): "Analysis of speech based on stochastic process model", Bull. Electrotechn. Lab. 39, 881-902.
- Ladefoged, P., R. Harshman, L. Goldstein, and L. Rice (1978): "Generating vocal tract shapes from formant frequencies", JASA 64, 1027-1035.
- Lindblom, B. and J. Sundberg (1969): "A quantitative model of vowel production and the distinctive features of Swedish vowels", STL-QPSR 1, 14-32 (KTH, Stockholm).
- Mrayati, M. and B. Guérin (1976): "Etude des caractéristiques acoustiques des voyelles orales françaises par simulation du conduit vocal avec pertes", Revue d'Acoustique No. 36, 18-32.
- Schroeder, M.R. (1967): "Determination of the geometry of the human vocal tract by acoustic measurements", JASA 41, 1002-1010.
- Wakita, H. (1973): "Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms", IEEE Trans. Audio and Electroacoustics AU-21, 417-427.
- Öhman, S.E.G. and S. Zetterlund (1974): "On symmetry in the vocal tract", in Speech Communication, Vol. 2, G. Fant (ed.), 133-138, Stockholm: Almqvist & Wiksell 1975 (Proc. SCS-74, Stockholm).

## MODERN METHODS OF INVESTIGATION IN SPEECH PRODUCTION

Osamu Fujimura, Bell Laboratories, Murray Hill, New Jersey 07974  
U.S.A.

Introduction

The natural process of speech production may be discussed on several levels beginning with the cortical level and ending with the acoustic signals. The higher the level, the less applicable direct physical measurements are. Recent efforts by psychologists are focused on temporal aspects of motor control, in an attempt to infer basic mechanisms of cortical programming and its execution. Techniques such as adaptation and reaction time measurements are now being used for direct observation of speech production processes (e.g. Sternberg et al. 1978), and it is hoped that such techniques in combination with powerful physical measurements of speech articulation processes will trigger a new development in this area of research.

Several interesting proposals have been made about the basic principle of articulatory dynamics trying to relate abstract and discrete phonological codes to the temporal structures of continuous speech phenomena (see Kent and Minifie 1977 for a review). The notion of coarticulation (Öhman 1967) still requires a general definition in relation to the basic process of concatenating well-defined phonetic units (Fujimura and Lovins 1978). Information on actual movements of the principal organs is badly needed for such a study. Relatively large amounts of data obtained from the same subject are necessary to cope with an inherent variability of speech production phenomena.

In what follows, we shall try to review recent work on physiological or physical (but not acoustic) observations. Due to the severe space limitation, reference can be made only to a small subset of the representative examples.

Physiological Studies - Muscle Controls

The general question here is which muscle plays the principal role of implementing motor commands for a given phonetic gesture, viz. an elementary articulatory event. Electromyographic studies with use of hooked-wire electrodes, for example by Hirose and Gay (1972), have revealed that the glottal abduction reflecting the devoicing gesture is related to activity of the posterior

cricoarytenoid muscles, whereas glottal adduction is achieved by several different muscles in varied ways depending on linguistic (and paralinguistic) functions.

Hirano (1977) recently studied the anatomy and physiology of the vocal cords using various advanced techniques such as electron-microscopy, histochemistry, electromyography, electric nerve stimulation, high speed motion picture, and mechanical measurements, applied to both human and animal larynxes. He arrived at a cover-body approximation of the vocal cords, which reminds us of the earlier account by Svend Smith. Baer has provided a detailed study of excised canine larynxes, and Titze and his coworkers are contributing a new computerized model of the vocal cord vibration process.

Lingual muscles are difficult to study, but the rather limited information obtained by EMG measurement is indispensable for inferring muscular functions relative to specific phonetic gestures. Of particular importance is the use of computational models simulating the tongue deformation as the result of muscular contraction patterns. A three-dimensional static model using the finite-element method has been initiated by Kiritani and substantially extended by Kakita. The role of orosensory patterns in defining targets of articulatory gestures has been discussed by Stevens and Perkell (1977) in relation to the quantal nature of speech. Controlled interference, by such techniques as anesthesia and bite block, has been used to study the effects of feedback on articulatory gestures. The complexity of speech physiology and the highly experienced human strategies in speech behavior tend to make an interpretation of the results of such experiments rather difficult, but some interesting findings are available (Lindblom et al. 1978). A servo-mechanistic technology for controlling mechanical load for dynamically specified load impedance can be used for a control-theoretical analysis of natural articulatory systems. According to Abbs and coworkers, the frequency response of feedback loop systems for articulators seems to allow actively controlled movements of articulatory organs via brainstem feedback, but there are occasions in speech articulations where so-called ballistic-type inertia-controlled movements of articulators are observed (Fujimura 1961). On the other hand, a dynamic palatographic study suggested feedback-controlled tongue tip movements for apical stop gestures (Fujimura

et al. 1973b).

#### Physical States of Organs

There are several stages of information mapping between physiologic motor control, the resultant muscular contraction patterns and the sound output signals. An efficient computational procedure for studying the relation between vocal tract area functions and formant patterns has been proposed by Mathews and coworkers (Atal et al. 1978). There is considerable interaction between the source and the vocal tract, and this situation can be computer-simulated by a composite vocal-cord vocal-tract system (Flanagan et al. 1975).

The physiologic control of the larynx is parametric in the sense that usually gross average states of the larynx rather than details of vibratory changes of the vocal cord shapes within each voice fundamental period are adjusted. The fiberscopic technique developed by Sawashima and Hirose (1968) or its stereoscopic version is appropriate for studying such parametric states of the larynx. Much knowledge has been gained by the use of the fiberscope. In particular, the state of the glottal aperture during the oral closure for stop consonants with various types of laryngeal control is now relatively well known for languages such as Korean, French, Hindi, Tibetan, as well as English, Swedish, and Japanese. Electric measurement of the glottal state is also useful for phonetic studies.

There have been several methods proposed and tested in the past decade for observing tongue movements: dynamic palatography, its extension to palato-lingual distance measurements, magnetic as well as ultrasonic measurements. The most direct and informative method for observation of tongue movement is the use of x-rays for lateral views of the tongue. There were two factors that made radiographic measurements impractical for obtaining a large quantity of speech data: radiological disturbance and the time-consuming frame-by-frame analysis. A new computer-controlled x-ray microbeam system was devised to overcome these technical difficulties (Fujimura et al. 1973a). A full-scale system is now in operation at the University of Tokyo (Kiritani et al. 1975), and is producing useful data about movements of metal pellets placed on selected points of the articulators. Computer programs have been designed and implemented at Bell Laboratories in order to give the experimenter an efficient tool for interactive data analysis. An auto-



matic algorithm has been devised which, according to specified phonetic symbols, identifies the time domains where relevant articulatory activities (and sound characteristics) are found. This system is useful both for assisting the experimenter in retrieving relevant parts of data, and for testing hypotheses about inherent characteristics of individual phonetic events.

An independent measurement of area functions by acoustic input impedance measurement has been proposed (Sondhi and Gopinath 1972).

#### Statistical Processing of Production Data

Through purely statistical processes, elementary component gestures of the tongue have been derived (Ladefoged 1977; Kiritani 1977). This inductive method gives us phenomenologically derived "phonetic coordinates" for describing articulatory characteristics of classes of phonetic units, classes defined by the particular choice of the speech material used for this data processing. How the results relate to our linguistic experience is an interesting question. Shirai and Honda (1977), along with other groups, assumed a simple dynamic model of the articulator movements to determine the parameters that characterize the natural system. These methods are useful particularly when we need preliminary guidelines for designing components of a larger-scaled deductive experimentation -- synthesis by rule with a comprehensive scope of simulation of human speech production.

#### References

- Atal, B. S., J. J. Chang, M. V. Mathews, and J. W. Tukey (1978): "Inversion of Articulatory-to-Acoustic Transformation in the Vocal Tract by a Computer-Sorting Technique", JASA 63, 1535-1555.
- Flanagan, J.L., K. Ishizaka, and K.L. Shipley (1975): "Synthesis of Speech from a Dynamic Model of the Vocal Cords and Vocal Tract", Bell Syst Tech J 54(3), 485-506.
- Fujimura, O. (1961): "Bilabial Stop and Nasal Consonants: A Motion Picture Study and its Acoustical Implications", JSHR 4, 233-247.
- Fujimura, O., S. Kiritani, and H. Ishida (1973a): "Computer-Controlled Radiography for Observation of Movements of Articulatory and Other Human Organs", Comput Biol Med 3, 371-384.
- Fujimura, O., I. F. Tatsumi, and R. Kagaya (1973b): "Computational Processing of Palatographic Patterns", JPh 1(1), 47-54.
- Fujimura, O. and J. B. Lovins (1978): "Syllables as Concatenative Phonetic Units", in Syllables and Segments, A. Bell and J. B. Hooper (eds.), North-Holland Pub. Co.
- Hirano, M. (1977): "Structure and Vibratory Behavior of the Vocal Folds", in Dynamic Aspects of Speech Production, M. Sawashima and F. S. Cooper (eds.), 13-30, U. Tokyo Press.
- Hirose, H. and T. Gay (1972): "The Activity of the Intrinsic Laryngeal Muscles in Voicing Control -- an Electromyographic Study", Phonetica 25, 140-164.
- Kent, R. D. and D. Minifie (1977): "Coarticulation in Recent Speech Production Models", JPh 5, 115-133.
- Kiritani, S., K. Itoh, and O. Fujimura (1975): "Tongue-Pellet Tracking by a Computer-Controlled X-ray Microbeam System", JASA 57(6,2), 1516-1520.
- Kiritani, S. (1977): "Articulatory Studies by the X-ray Microbeam System", in Dynamic Aspects of Speech Production, M. Sawashima and F. S. Cooper (eds.), 171-190, U. Tokyo Press.
- Ladefoged, P. N. (1977): "The Description of Tongue Shapes", in Dynamic Aspects of Speech Production, M. Sawashima and F. S. Cooper (eds.), 209-222, U. Tokyo Press.
- Lindblom, B., R. McAllister, and J. Lubker (1978): "Compensatory Articulation and the Modeling of Normal Speech Production Behavior", in Articulatory Modeling and Phonetics, R. Carré, R. Descout, and M. Wajskop (eds.), 148-161, G.A.L.F. Groupe de la Communication Parlée.
- Öhman, S. E. G. (1967): "Numerical Model of Coarticulation", JASA 41(2), 310-320.
- Sawashima, M. and H. Hirose (1968): "New Laryngoscopic Technique by Use of Fiber Optics", JASA 43(1), 168-169.
- Shirai, K. and M. Honda (1977): "Estimation of Articulatory Motion", in Dynamic Aspects of Speech Production, M. Sawashima and F. S. Cooper (eds.), 279-302, U. Tokyo Press.
- Sondhi, M. M. and B. Gopinath (1972): "Determination of Vocal-Tract Shape from Impulse Response at the Lips", JASA 49, 1867-1873.
- Sternberg, S., S. Monsell, R. L. Knoll, and C. E. Wright (1978): "The Latency and Duration of Rapid Movement Sequences: Comparisons of Speech and Typewriting", in Information Processing in Motor Control and Learning, G. E. Stelmach (ed.), 117-152, Academic Press.
- Stevens, K. N. and J. S. Perkell (1977): "Speech Physiology and Phonetic Features", in Dynamic Aspects of Speech Production, M. Sawashima and F. S. Cooper (eds.), 323-341, U. Tokyo Press.

THE PHYSIOLOGY AND PATHOPHYSIOLOGY OF LANGUAGE FUNCTIONS AS  
ILLUSTRATED BY MEASUREMENTS OF THE REGIONAL BLOOD FLOW IN THE  
CORTEX OF THE BRAIN

Niels A. Lassen, Department of Clinical Physiology, Bispebjerg  
Hospital, Copenhagen

By measuring the blood flow in small regions of the brain (Xenon-133 injection via the internal carotid artery), an increase in flow is seen that corresponds to an increase in metabolism and neuronal function in the same region (Lassen et al. 1978). Typically the regional flow increases by 30%. Simple sensory perception or motor performance activate the well known respective primary and secondary areas.

When one listens to speech, is speaking oneself, or reads aloud, then 2, 3, respectively 6 regions become simultaneously active in both hemispheres (minor side-to-side differences appear to exist, and will be commented upon).

When listening to words, the 2 active areas are I: The temporal lobe, superior-posterior part (on the left side, comprising Wernicke's posterior speech center), II: an inconstant activation over the inferior frontal region (on the left side, comprising Broca's anterior speech center). This region overlies the basal ganglia and hence it cannot be decided if this area is cortical or subcortical. The inconstancy of the activation could mean that even at rest, the thought processes involve (inconstantly?) an activity in this area.

When speaking, the 3 active areas are I: The temporal lobe (see above), II: The primary mouth area in the central region, III: The supplementary motor area high in the frontal lobe (Penfield's superior speech area). In automatic speech in the form of counting to twenty repeatedly, we see little hyperactivity in the lower frontal area. But in fluent normal speech this area is very often active.

When reading a simple text aloud, the 6 active areas are: I: The temporal lobe (comprising on the left side Wernicke's area), II: usually the inferior frontal region (probably comprising Broca's area on the left side), III: The motor mouth area, IV: The supplementary motor area (Penfield's superior speech area), V: The visual association cortex, VI: The frontal eye field.

All these changes are bilateral. We do not see the primary visual cortex (it is supplied from the vertebral artery). But animal studies clearly show this area also to become activated during visual perception. Hence, during reading, even with the coarse resolution ( $1 \text{ cm}^2$ ) of our method, reading can be said to involve the collaboration of (at least) 14 discrete cortical areas, 7 in each hemisphere.

The lecture will comment on the differences between listening to noise and to words. Consideration as to the changes in aphasia as well as to the possible contribution of the right hemisphere to language functions ("emotional colour", prosody) will also be discussed.

#### References

Lassen, N.A., D.H. Ingvar, and E. Skinhøj (1978): "Brain function and blood flow", Scientific American 239, 62-71.



## NEW METHODS OF ANALYSIS IN SPEECH ACOUSTICS

Hisashi Wakita, Speech Communications Research Laboratory, Inc.,  
806 West Adams Boulevard, Los Angeles, California 90007, U.S.A.

Introduction

The recent development in digital techniques has brought substantial innovations to methods and techniques for acoustical analysis of speech sounds. The advantages of using digital computers over the conventional analog techniques are that the analysis processes can be repeated precisely and that the control of parameters is relatively easy. The use of a digital computer also permits the processing of a large amount of data within a relatively short period of time with satisfactory accuracy. Because of the above advantages, digital techniques are playing a more and more important role in speech research. This paper, thus, concerns primarily the recent digital techniques in speech analysis, particularly the linear prediction method, with special attention to its advantages and disadvantages, and also the limitations involved in the techniques.

Analysis Techniques

Among the current digital techniques for speech analysis, the linear prediction method (LP method) is predominantly used by many researchers. The LP method is suited for digitally processing the speech data to extract some acoustic parameters (Wakita, 1976).

## a) Formant Analysis

The LP method assumes a simple speech production model

which consists of an excitation source and a transmission system. The excitation source is assumed to be an impulse generator, and thus the transmission system includes a glottal shaping filter, a vocal tract filter, and a lip radiation filter. Since the vocal tract filter does not assume a nasal tract, the LP model primarily assumes the production of voiced nonnasal sounds. Analysis by the LP method is an attempt to match a speech segment to the above ideal model so that the error between the matched model and the ideal one becomes minimum on a least mean square error criterion. The transmission system, thus optimally determined, is represented by either a set of predictive coefficients or a set of reflection coefficients. A smooth spectral envelope can be obtained by applying a Fourier transform to a given set of predictive coefficients. The formant frequencies can be obtained either by searching for peaks in the spectral envelope or by precisely computing the roots of the polynomial of predictive coefficients. By the above procedure, the formant frequencies are fairly accurately estimated. The formant bandwidth or amplitude can also be computed, but its accuracy is sometimes erroneous. Like other methods for estimating formant frequencies, the LP method does not solve the problem of estimating formant frequencies for speech sounds of very high pitch. This problem rather inherently exists in the speech signal, and thus it is rather intrinsic to any method.

#### b) Detection of Fundamental Frequency

In the LP analysis, after the vocal tract characteristics are extracted from the speech signal, the information on the fundamental frequency still remains in the residual. The

periodicity of the speech signal can thus be extracted from the residual signal by the autocorrelation method. Besides the LP method, there are various methods for extracting the fundamental frequency. Most of them have a high performance and have their advantages and disadvantages (Rabiner et al., 1976). The choice of a method depends upon the purpose, speakers, and recording conditions.

#### c) Other Topics

The reflection coefficients obtained by the LP analysis have been shown to give an acoustic tube representation of the transmission system in the LP model. Thus, if some appropriate preprocessing is applied to a speech segment to eliminate the excitation source characteristics and the effect of the lip radiation load, a realistic area function of the vocal tract is expected to be recovered from acoustic analysis of the speech waves (Wakita, 1979). Although the precise determination of the vocal tract shapes by acoustic analysis of the speech waves is difficult, a fairly good approximation to them is expected.

In an attempt to estimate the glottal characteristics, the actual vocal tract characteristics can also be estimated from a portion of a voiced sound during which the glottis is closed. After the vocal tract characteristics thus estimated are eliminated from the speech signal, the glottal volume velocity waves are recovered from the residual signal.

This trend of extracting some articulatory parameters from the speech waves stimulated the development of other types of articulatory models to which the speech signal is directly transformed (Atal, 1974).

Conclusion

Based on the above acoustic analysis of speech sounds, the LP method has various potential applications to many areas of speech research. The method will be a powerful tool to investigate the interrelationships between articulation and its acoustic characteristics with the aid of, results from other direct physiological measurements. This would contribute to a more complete articulatory model for understanding speech production as well as to a better speech synthesizer. Application of the techniques to speech feature extraction and segmentation will eventually make the automatic transcription of speech sounds possible.

References

- Atal, B. S. (1974): "Towards determining articulator positions from the speech signal," Preprints of the 1974 Stockholm Speech Communication Seminar, Vol. 1, 1-9.
- Rabiner, L. R., M. J. Cheng, A. E. Rosenberg, and C. A. McGonegal (1976): "A comparative performance study of several pitch detection algorithms," IEEE ASSP, ASSP-24, 5, 399-418.
- Wakita, H. (1976): "Instrumentation for the study of speech acoustics," in Contemporary Issues in Experimental Phonetics, N. J. Lass (ed.), 3-40, New York: Academic Press.
- Wakita, H. (1979): "Estimation of vocal tract shapes from acoustical analysis of the speech waves: its status of the art," IEEE ASSP (to be published).

## VOYELLES LABIALES ET VOYELLES LABIALISEES EN FRANCAIS

## Etude labiographique

Christian Abry, Louis-Jean Boë, Institut de Phonétique de Grenoble, Raymond Descout, C.N.E.T., Lannion

L'arrondissement en français est un trait de mode à la fois pour les voyelles [y, ø .../ i, e ...] et les consonnes [ʃ, ʒ.../ s, z ...]. Nous avons étudié le comportement de ce trait dans le cas où du fait des règles de coarticulation, l'arrondissement des consonnes, qui est non phonologique, assimile les voyelles phonologiquement arrondies.

Le corpus est constitué de mots où figurent, en position finale accentuée, les syllabes CV avec C = s, z, ʃ, ʒ et V = i, e, y, ø, soit 96 réalisations. Les mots ont été placés dans des phrases porteuses. Un labiofilm, face et profil (35mm, 50 images/s., son synchrone) a été réalisé pour 5 locuteurs (2 femmes, 3 hommes). Les paramètres retenus sont: l'écartement intérolabial (A), l'aperture entre les lèvres (B), l'aire intérolabiale (S), l'aperture extérolabiale à l'extrémité du conduit vocal, la protrusion-rétraction des lèvres supérieure et inférieure ( $F_1$ ,  $F_2$ ), la position de la mâchoire (M), le point de contact des lèvres (C), la distance entre C et la tangente  $F_1 F_2$  (L). Les images projetées par agrandisseur, sont acquises par ordinateur grâce à une tablette d'entrée graphique, puis traitées statistiquement.

L'examen de la distribution des données montre que:

- Les valeurs + et - du trait [rond] sont de nature différente: les voyelles phonologiquement [+rond] tendent vers une constante de forme A/B, les voyelles [-rond] vers une constante d'aire  $A \times B$ .
- La frontière phonémique n'est pas obtenue, contre toute attente, avec  $F_1$ ,  $F_2$ , mais avec A et B, les paramètres les plus significatifs. Elle peut être optimisée avec une valeur de l'aire aux lèvres. - L'assimilation consonantique ne met en évidence, toujours avec A et B, qu'une seule classe subphonémique [i,e].
- La frontière phonémique est fragile du seul point de vue des paramètres articulatoires, mais il est probable qu'elle correspond à une limite naturelle, une frontière phonétique entre deux quanta (Stevens, 1972).

#### Référence

Stevens, K.N. (1972): "The quantal nature of speech: evidence from articulatory-acoustic data", in E.E. David and P.B.Denes (eds.) Human communication: a unified view, 51-66, New York: McGraw Hill.

ETUDE ELECTROPALATOGRAPHIQUE ET AERODYNAMIQUE SIMULTANEE  
DES OCCLUSIVES FRANCAISES

D. Autesserre et B. Teston, Institut de Phonétique, Aix-en-Provence

Deux prototypes d'électropalato-graphes, ELPA II et ELPA III, mis au point à l'Institut de Phonétique d'Aix-en-Provence, sont utilisés pour recueillir et analyser les contacts successifs des diverses parties de la langue contre le palais pendant la réalisation des consonnes occlusives du français. Simultanément, on évalue les variations de débit d'air au sortir de la cavité buccale à l'aide du polyphonomètre. La combinaison de ces deux types d'investigation instrumentale conduit le chercheur en électropalato-graphie à prendre conscience des difficultés de détection des contacts linguaux. Certaines électrodes palatines situées dans la région d'occlusion, ne sont pas activées comme on aurait pu s'y attendre. Toutefois, le caractère systématique de certains décalages entre l'interruption du flux d'air et le déclenchement de l'inscription des électrodes palatines concernées nous emmène à envisager sous un angle nouveau les conditions d'installation de l'occlusion linguale. Par ce type d'exploration on a pu déterminer:

1) l'ordre de succession des appuis linguaux par rapport à la période d'interruption de l'air expiré;

2) la durée relative des contacts de la langue aux divers endroits du palais par rapport à la durée globale des trois phases successives de la production d'une occlusive (catastase, tenue, métastase).

Les résultats obtenus viennent compléter les renseignements fournis par d'autres techniques (radiocinéma, électromyographie) et apportent une contribution indispensable à l'étude de la dynamique linguale.

Références

Autesserre, D. et B. Teston (1978): "Description of a dynamic electropalatographic system", Proceedings of the Phonetic Sciences Congress IPS-77, Miami (sous presse).

Teston, B. (1976): "Description d'un système d'analyse des paramètres articulatoires", Travaux de l'Institut de Phonétique d'Aix 3, 151-207.

## ACOUSTIC CORRELATES OF DIFFERING ARTICULATORY STRATEGIES

F. Bell-Berti<sup>1</sup>, L. J. Raphael<sup>2</sup>, D. B. Pisoni<sup>3</sup>, and J. R. Sawusch<sup>4</sup>,  
Haskins Laboratories, New Haven, Connecticut, U. S. A.

In an earlier EMG and vowel identification study (Bell-Berti et al., 1978) we hypothesized that inter-speaker differences in the perception of vowels, variously described as differing in tongue tension or tongue height, reflect differing articulatory strategies. In an attempt to explain the perceptuo-productive relationships more fully, we subjected the utterances of the nine subjects in the EMG experiments to acoustic analysis.

Method

The acoustic analyses were performed on /əpVp/ utterances, where V=/i,ɪ,e/ or /ε/, using a digital waveform and spectral analysis system. Averages of the first three formant frequencies and vowel durations, for each speaker, were computed for a minimum of 15 repetitions of each utterance.

Results

Preliminary results do not reveal systematic differences between formant frequency patterns of speakers using differing articulatory strategies. Inter-speaker differences, however, were revealed by durational analysis: speakers who differentiated the members of the /i-ɪ/ and /e-ε/ pairs on the basis of tongue tension showed a greater durational difference between the members of the pairs than did speakers who used tongue height to differentiate the members of the pairs.

References

Bell-Berti, F., L. J. Raphael, D. B. Pisoni, and J. R. Sawusch (1978): "Some relationships between articulation and perception," Haskins Laboratories Status Report on Speech Research SR-55.

- 
- (1) Also, Montclair State College, Upper Montclair, New Jersey (on leave)  
 (2) Also, Herbert H. Lehman College, City University of New York, New York, New York  
 (3) Indiana University, Bloomington, Indiana  
 (4) State University of New York at Buffalo, Buffalo, New York

ETUDE EXPERIMENTALE DE CERTAINS ASPECTS DE LA GEMINATION ET DE  
L'EMPHASE EN ARABE

Jean-François P. Bonnot, Instituut voor Romanistiek, Universit    
d'Amsterdam, Singel 134, 1015 AG Amsterdam, Pays-Bas.

A partir d'un corpus d'arabe classique, nous avons   tudi    
les manifestations de l'emphase et de la g  mination. L'entourage  
des occlusives /t tt t tt/ se limite    la voyelle /a/. Nos r  sul-  
tats proviennent de 2 radiofilms (42 images/sec), d'oscillogrammes  
   4 lignes et de sonagrammes en filtrage large. Les films ont   t    
mesur  s gr  ce    la m  thode des axes, trac  s de 20 en 20 degr  s.

R  sultats et discussion

Le rapport des dur  es g  min  e/simple est de 1.76 pour les  
emphatiques et de 2.02 pour les non-emphatiques. Les g  min  es pr  -  
sentent un renforcement de l'articulation et une accentuation des  
propri  t  s des consonnes simples: pour /tt/, fermeture et ant  rio-  
risation; pour /tt/, augmentation du caract  re emphatique (exten-  
sion de la constriction post  rieure en direction de la zone uvu-  
laire). On ne rel  ve aucun mouvement sp  cifique de la paroi post  -  
rieure du pharynx, ni pour la simple - nous sommes en accord avec  
Ali et Daniloff (1972) - ni pour la g  min  e. L'indice de dispersion  
articulatoire (langue et maxillaire) est moins important pour le  
degr   long de quantit  , qu'il s'agisse des emphatiques ou des non  
emphatiques. Quant aux voyelles, leur articulation diff  re selon  
qu'elles sont au contact d'une simple ou d'une g  min  e, et de ce  
fait, leur structure acoustique est soumise    quelques variations.

Enfin, les donn  es dont nous disposons, et particuli  rement  
l'examen image par image des radiofilms, ne montrent aucun signe  
de r  articulation (Lehiste, 1973) de la g  min  e. Non seulement on  
ne constate aucune marque de fl  chissement durant la tenue, mais  
de plus la stabilit   organique est tr  s sup  rieure. N  anmoins, les  
configurations articulatoires diff  rentes et la faible dispersion  
des mesures semblent r  v  ler une programmation sp  cifique des con-  
tr  les moteurs pour les g  min  es et leur entourage imm  diat.

R  f  rences

- Ali, L.H. et R.G. Daniloff (1972): "A cinefluorographic investiga-  
tion of emphatic sound assimilation", Proc.Phon. 7, 639-648.  
Lehiste, I., K. Morton, et M.A.A. Tatham (1973): "An instrumental  
study of consonant gemination", JPh 1, 131-148.



FEEDBACK AND FEEDFORWARD MECHANISMS USED BY SPEAKERS PRODUCING  
FAMILIAR AND NOVEL SPEECH PATTERNS

G.J. Borden, K.S. Harris, H. Yoshioka, and H. Fitch

Haskins Laboratories, New Haven, Connecticut U. S. A.

Speech has been shown to be remarkably stable despite attempts to interrupt sensory feedback. The present study indicates that speech production control operates somewhat differently when the task involves imitation of unfamiliar utterances. Data were collected on two normal subjects imitating a phonetician producing syllable patterns. Some of the syllables were familiar to the subjects /pi/, pe<sup>t</sup>/, /fi/, and /zi/, while others were less familiar /py/, /pø/, /xi/, and /yi/. Subjects repeated the imitations under various combinations of abnormal speaking conditions: nerve block anesthesia, auditory masking, and an artificial extension of the alveolar ridge.

Analysis of the data includes sound spectrograms, EMG recordings of pertinent articulatory muscles, and a test made for listener judgments of the imitations produced under the various conditions.

Results from the first subject show novel utterances to vary more than familiar utterances in vocal tract resonances and in EMG patterns. When the vocal tract area alteration was added to the nerve block plus masking condition, listeners judged the imitations to be worse, as speakers are presumably forced to change positional goals to come close to their auditory perceptual goals. The condition in which the speaker could hear himself despite loss of tactile sensation resulted in higher front cavity resonances and more accurate imitations, indicating that self hearing sharpens the match between vocal tract shape and perceptual goals. Results of the study will be interpreted within the framework of a model of speech production regulation, which operates differently for speech acquisition than for production of skilled speech.

References

- Borden, G.J., K.S. Harris, and L. Catena (1973): "Oral feedback II. An electromyographic study of speech under nerve-block anesthesia", JPh 1, 297-308.
- Borden, G.J. (1976): "The effect of mandibular nerve block upon the speech of four-year-old boys", L&S 19, 173-178.

## MICROTIMING OF TWO-CONSONANT CLUSTERS

Blanka Borovičková and Vlastislav Maláč, Academy of Sciences and Institute of Radiocommunication, Prague, Czechoslovakia

By the term microtiming differences of sound duration caused by the coarticulation processes are meant.

Subject

Clusters involving 80 consonants grouped into symmetrical VCCV sound combinations were analysed from this point of view. All differences in duration were expressed in a logarithmic scale, decichron, which is defined as  $dC = 10 \log T_0/T$ , where  $T_0$  is the average duration of a sound and  $T$  is the duration of the sound measured. It was found that the durational differences depend on 1) the position within the cluster (first or second), 2) the kind, 3) the environment of a given consonant. These influences were summed up in two equations.

The difference of the first consonant in a cluster is given by the equation:

$$\Delta C_1 = k_{nC1} + 0,2 + (n_{C2} - 3) \cdot 0,3 \quad /dC/ ,$$

where  $k$  is a coefficient which indicates a durational difference (in  $dC$ ) for one of the five classes of consonants marked  $n$ ; the second member of this equation represents the average extension of the first consonant duration; the influence of the second consonant is expressed by the third member of the equation. The difference of the second consonant is given by the equation:

$$\Delta C_2 = k_{nC2} - 0,7 + (3 - n_{C2}) \cdot 0,3 + (n_{C1} - 3) \cdot 0,3 \quad /dC/ ,$$

where the first member represents a correction of the consonant class; the second means the average shortening of the second consonant; the third member expresses the equalization tendency of the second consonant in the class differences (represented by the  $k$  coefficient); the last member represents again the influence of the first consonant on the second one.

Conclusion

Only 2,5% of the calculated consonants' durational difference in the two-consonant clusters deviate from the measured ones, within a set of 80 consonant clusters. We considered the perceptually significant differences only; i.e. greater values than 1  $dC$  (roughly 20%).

## ARTICULATORY TIMING IN VOICELESS FRICATIVES

Andrew Butcher, Department of Linguistic Science, University of Reading, U.K.

This paper presents some airflow data on the production of intervocalic voiceless fricatives by German speakers, points out some interesting features, and suggests a (micro-)model to account for them.

The speech sounds concerned ([f, s, ʃ, x] and later also [ç]) originally occurred in the word final position of verbs spoken in a frame at two speeds and with three variations in sentence stress. The data consisted of measurements made from airflow and laryngograph traces of four native speakers, later supplemented by curves from VCV sequences pronounced by two phoneticians. The majority of traces exhibit the characteristic twin peaks of airflow observed by other investigators, some of whom have also offered the explanation that these are the result of timing differences between glottal and supraglottal articulation (Klatt et al., 1968, 48). Amplitudes for apicals and labials are in general lower than those for tongue body fricatives. With the latter it was found that peaks immediately adjacent to homorganic vowels (i.e. [ç + i] and [x + u]), if present at all, are lower than those next to a vowel requiring a different tongue position - the highest peaks being those adjacent to [a].

In other words, it seems that the longer the tongue has to move from one segment to the next, the higher the rate of airflow reached at the transition. A possible explanation for this might be that the motor commands for the movements of the supraglottal articulations are given at a fixed point in time relative to those for the abduction and adduction of the vocal folds, so that the discrepancies in the timing of these actions, and hence the amplitudes of the airflow peaks, would be to a great extent a function of the difference in place of articulation for vowel and consonant.

Reference

Klatt, D.H., K.N. Stevens and J. Mead (1968): "Studies of articulatory activity and air flow during speech", Annals of NY Academy of Sciences 155, 42-55.

## THE ARTICULATORY FUNCTION OF THE VELUM

Luiz Carlos Cagliari, Departamento de Lingüística  
UNICAMP, IEL, Caixa Postal 1170, 13.100 - Campinas, Brasil

The function of the velum in the production of speech has been investigated since the XVIIIth century. The main aspects recently investigated are the anatomy of the region, the muscular action, the oral-nasal feature in languages, the acoustics of nasality and the interaction of the velic action with other parameters in the production of speech (Cagliari, 1977).

In the description of phonetic segments in languages, it is common to incorporate only two velic positions: the elevated velum in the production of oral segments, and the lowered velum in the production of nasalized segments. However, instrumental investigations have shown that velum assumes different positions as a function of different phonetic segments. The reason for this is the inherent susceptibility of these segments to nasalization and perhaps neuromuscular constraints associated with the functioning of other articulators. For this reason, it seemed interesting to suggest an articulatory model of the velum based on a neutral velic scale. Acoustic and EMG investigations, as well as perceptual tests, have corroborated this hypothesis. The suggested model of velic action gives a better understanding of the nature of nasality and denasality as two types of voice quality, of the relation between the segmental features of nasality and orality linguistically. Finally, it shows more precisely how different degrees of nasality and denasality are performed.

Reference

Cagliari, L.C. (1977) An Experimental Study of Nasality with Particular Reference to Brazilian Portuguese, unpublished Ph.D. Thesis, University of Edinburgh.

TOMOGRAPHIC REGISTRATION OF THE FRONT ORAL CAVITY AT THE  
PRONUNCIATION OF S

Olof Eckerdal, Dept. of Oral Roentgenology, and Claes-Christian Elert, Dept. of Phonetics, Umeå University, Umeå, Sweden

The pronunciation of [s] of 22 Swedish-speaking subjects was studied on three consecutive tomographic frontal layers of the molar, premolar and cuspid regions of each subject. Tomographic roentgenograms in frontal projection of the foremost part of the [s]-channel cannot be taken successfully because of the steepness of the palate curvature close to the incisors (Eckerdal 1973). Xerographic technique was used, allowing registration of soft-tissue as well as hard-tissue contours (Schertel, 1975). For 10 of the subjects the soft-tissue contours were checked on molds.

According to traditional phonetic theory, an [s] is produced with a longitudinal tongue groove. The predorsal position prevails among Swedish speakers. The radiographic images gave data towards a specification of the tubular cavity formed by the tongue, the teeth and the roof of the mouth. The cross-section areas at the three layer positions were calculated. The shape of the bottom of the groove was rounded in most subjects. It had a narrow, V-like outline in about 25% of the cases. There was a deviation of the groove from the midline of the oral cavity in about 80% of the subjects, mostly to the right (50%).

References

- Eckerdal, O. (1973): "Tomography of the temporomandibular joint", Acta Radiologica, Suppl. 329.
- Schertel, L. (1975): "Die Anwendung des xerographischen Verfahrens bei der Tomographie von Schädel und Hals", Fortschritte auf dem Gebiet der Röntgenstrahlen und der Nuklearmedizin 122, 295-300.

## AERODYNAMIC MEASUREMENTS ON ITALIAN INTERCONSONANTAL VOWELS

Edda Farnetani, Centro di Studio per le Ricerche di Fonetica, Padova, and

Jan Gauffin, Dept. of Speech Communication, KTH, Stockholm

The aim of this paper is to show that air flow measurements, controlled by air pressure measurements, may be adequate for a qualitative description of articulatory dynamics. In particular, the application of this technique to a speaker of Standard North Italian during production of short sentences has confirmed the results of a previous acoustic analysis (Vagges et al. 1975) about vowel length variations due to the following voiceless/voiced stops and has made it possible to correlate acoustic length variation to different movements of the articulatory structures. It was found that acoustic differences in durations reflect in part different speed of movements and are in part the result of different glottal adjustments. This study has also made it possible to correlate the presence of the "voice bar" in both single and geminate intervocalic voiced stops to an active expansion of the supraglottal cavity, which starts during the preceding vowel (30-40 ms before the closure). This movement seems to be quite independent of the closing gesture, which is taking place at the same time for both bilabial and dentoalveolar voiced stops, but seems to interfere with the closing gesture for velar voiced stops.

References

- Ohala, J.J. (1974): "A mathematical model of speech aerodynamics", Speech Communication Seminar Stockholm, 2, 65-72.
- Rothenberg, M. (1968): The Breath-Stream Dynamics of Simple-Released Plosive Production, Bibliotheca Phonetica No. 6.
- Rothenberg, M. (1977): "Measurement of airflow in speech", JSHR 20, 155-176.
- Vagges, K., F.E. Ferrero, E. Caldognetto-Magno, and C. Lavagnoli (1975): "Some acoustic characteristics of Italian consonants", 8th Intern. Congress of Phonetic Sciences.

ELEKTROPALATOGRAPHIE ALS KONTROLLHILFE FÜR DAS ARTIKULATIONS-  
TRAINING IM GEHÖRLOSENUNTERRICHT

Slavko Geršič, Institut für Phonetik der Universität Köln

Dirk Steffen Schröder, Zahn- und Kieferklinik der Universität Köln

Wir berichten von den Arbeiten über das Problem der Elektropalatographie, die raum-zeitliche Echtzeitinformationen über Zungen-Gaumen-Kontakte während des Sprechens liefert. Das primäre Interesse liegt darin, Apparaturen zu entwickeln, die es dem Gehörlosen beim Versuch, die Sprechfähigkeit zu erwerben, ermöglichen, auf einem Fernschirmschirm sichtbar zu machen, was bei seinen eigenen Lautbildungsversuchen in seinem Mund geschieht, diese Produktionen mit Mustern zu vergleichen und auf diese Weise die eigene Lautproduktion zu verbessern. Als Aufgabe für die Zukunft ergibt sich aus der jetzigen Arbeit die Suche nach Wegen, wie die künstlichen Gaumen - die ja jeweils individuell angepaßt werden müssen - mit geringerem technischen und finanziellen Aufwand als bisher konstruiert werden könnten. Dies gilt auch für das von M. Lexa in unserem Institut entwickelte Interface. Was den künstlichen Gaumen betrifft, so wollen wir auf die gedruckten Schaltungen übergehen, was den obigen Anforderungen weitgehendst entgegenkommen würde. Die Frage nach einem vertretbar preiswerten Interface ist bereits gelöst.



MOUTH SHAPE IN THE PRODUCTION OF [w] AND [ɸ] SOUNDS IN JAPANESE  
Shizuo Hiki, Research Institute of Electrical Communication,  
Tohoku University, Sendai, Japan, and Yumiko Fukuda, Faculty of  
Education, Tohoku University, Sendai, Japan

The mouth of a speaker was illuminated by a stroboscopic light source every 10 milliseconds, and pictures of both frontal and lateral views of the mouth were taken utilizing a special camera in which a long film was driven continuously. Changes in the dimensions of various parts of the mouth were measured.

The up-and-down movements of the centers of the upper and lower lips, and the lateral movements of the corners of the lips, were also recorded by attaching small metal pellets to the lips at these points and by illuminating them with the stroboscopic lamp every 5 milliseconds. The frontal projections of the traces of these points were displayed three-dimensionally by adding the time axis. (This graph is called a "labiogram".)

The material used here comprised the traditional 100 Japanese monosyllables and some additional syllables occurring in loan words in modern Japanese. Some of the latter words consisted of two, three or four syllables. The words were spoken by a female adult.

On the basis of the stroboscopic observations and a spectrographic analysis of the speech sounds, characteristics of the mouth shape for each of the syllables and coarticulation effects were analyzed.

Among the results, this paper will focus on the characteristics of the sound [w] and the unvoiced bilabial fricative [ɸ], which are pronounced frequently in loan words, as well as in the traditional Japanese syllables /'wa/ and /hu/. (/'/ is the voiced counterpart of /h/.)

The use of visual information on the mouth shape for these sounds to improve lipreading of modern Japanese, will also be discussed.

## STRUCTURE OF THE VOCAL FOLD AS A SOUND GENERATOR

Minoru Hirano, Shigenobu Mihashi, Takao Kasuya and Shigejiro Kurita, Department of Otolaryngology, Kurume University, Kurume, Japan

The vocal folds, i.e. the sound generator, participate in differentiating voiced and voiceless speech sounds and in determining prosodic characteristics in speech. One single pair of vocal folds can cover great varieties of fundamental frequency and tonal qualities. This indicates that the vocal folds can become vibrators with many different mechanical properties. This paper presents some important aspects of the structure of the vocal folds which is adequate to the task.

Light and electron microscopic observations were conducted with human vocal folds. In addition, networks of the blood vessels of the vocal folds were investigated with an X-ray technique.

Histologically, the vocal folds consist of the mucosa and the muscle. The mucosa, in turn, consists of the epithelium and the lamina propria. The lamina propria has three layers: the superficial layer which is loose in fibrous component, the intermediate layer which is chiefly composed of elastic fibers, and the deep layer which is dense with collagenous fibers. From a mechanical point of view, we differentiate the layers into three sections: the cover consisting of the epithelium and the superficial layer of the lamina propria, the transition consisting of the intermediate and deep layer of the lamina propria, and the body, consisting of the vocalis muscle. The transition appears to be more closely connected to the body than to the cover as far as the histological evidence reveals. Based on the evidence of the networks of the blood vessels, the transition is more closely connected to the cover than to the body.

The cover and transition receives only passive adjustment, whereas the body is a subject to active and passive control.

## MECHANISMS FOR THE CONTROL OF VOCAL FREQUENCY

Harry Hollien and James W. Hicks, Jr., IASCP, University of Florida, Gainesville, Fl. USA

It is well known that control of vocal frequency is determined primarily by variation in the mass and stiffness of the vocal folds and by subglottic pressure. In turn, variation of vocal fold mass and stiffness results (in part) from changes in vocal fold length. But how is vocal fold lengthening accomplished? There is no question but that the major control results from contraction of the cricothyroid muscle which reduces the CT space and, hence, stretches the vocal folds. However, this poster presentation challenges the notion that this, the CT, action is the only major factor in that regard.

In order to investigate these relationships, the following steps were taken. 1) Mean and maximum variations in vocal fold length were calculated from appropriate research reports. 2) Estimates of laryngeal cartilage size were obtained from the literature. 3) A variety of potential laryngeal cartilage dimensions were calculated. 4) Based on these values, a computer program was run that tested the effect of the CT mechanism on vocal fold length. It was found that the CT activity could not account for all of the lengthening observed. A complementary mechanism then was sought.

From examination of lateral radiographs, it has been observed that the shadows of the arytenoids appear to move posteriorly as a function of increases in  $F_0$ . This apparent movement was measured on the X-rays of a number of subjects. The values obtained appear to account for the balance of vocal fold elongation. The results of some EMG studies support this explanation, the results of others do not.

## MECHANISMS OF ADOLESCENT VOICE CHANGE

Patricia A. Hollien and E. Thomas Doherty, IASCP,  
University of Florida, Gainesville, FL USA

In the past, a large number of studies have been carried out investigating adolescent voice change (AVC) and puberty. Most phoneticians who have studied these processes have tended to concentrate primarily upon voice features -- only occasionally including other variables. Further, most investigators have utilized relatively small populations and have employed a cross-sectional approach; hence, they have found it virtually impossible to provide information for individual subjects relative to any pubescent related process. Finally, most investigators have utilized primarily descriptive approaches when attempting to define and explain the nature of adolescent voice change -- and puberty.

In this investigation, attempts are made to meet these problems. The data-base was obtained from a relatively large group of males (N=48) studied longitudinally at bi-monthly intervals for a period of over four years. Fourteen variables were studied: age, five voice parameters and eight body dimensions. Finally, a cluster analysis statistical technique was utilized in order to permit 1) pre-, neo- and post-adolescent categories to be generated, 2) both group and individual pubescent status to be identified and 3) the classifications to be compared to those developed by traditional methods.

The cluster approach provided the three expected categories with the neo-adolescent group remaining stable and robust no matter how many clusters were specified. Tentative group predictions were established for American youths, and the status of a new subject can be determined by comparing his data to the various categories. Finally, when this method was compared to a traditional category approach, a higher mean neo-adolescent speaking fundamental frequency (SFF) (217 vs 178 Hz) was found but mean age was somewhat lower than that previously specified (13.5 vs 14.3 years). It can be concluded that the approach here utilized is useful as it permits the pubescent categories to be specified on the basis of easily applied group means and variabilities.

## NEUROMECHANICAL COMPONENTS OF REACTION TIMES FOR VOICE INITIATION

K. Izdebski and T. Shipp, Voice Science Laboratory, Department of Otolaryngology, University of California San Francisco, California, and Speech Research Laboratory, Veterans Administration Hospital, San Francisco, California, USA

This study investigated basic human sensory-motor processes underlying voluntary reaction time (RT) latencies for voice initiation through simultaneous aerodynamic, acoustic and electromyographic (EMG) recordings. Four adult subjects were pretrained to respond as quickly as possible following an auditory (1000 Hz sine wave) or a somesthetic (6 cm H<sub>2</sub>O of intraoral air pressure release) stimulus. Each subject provided data on neuromechanical reaction time latency that is the period from the stimulus onset to the initiation of vocal response. The component of RT comprising mechanical time (MT), in this case the latency from the onset of the interarytenoid and or posterior cricoarytenoid muscle activity to the initiation of the vocal response was measured directly. When the MT component is subtracted from the RT, it yields values for neural time (NT), that corresponds to the latency between the stimulus onset and the onset of the EMG activity. Neural time component is subdivided into the three sequential stages of afferent, cortical and efferent time. These three neural stages are time estimated. All RT components are discussed with reference to the stimuli used, and the phonatory task accomplished. The NT component was shown to be variable while the MT was shown to be stable independently of stimulus type and or the overall RT latency. Implications of RT variability in reference to more complex phonemic-linguistic load are discussed. (Supported by UCSF MSF Grant No 16 and VA's MRS.)

## AUDITORY FEEDBACK AS A FACTOR IN DISRUPTED SPEECH PRODUCTION

Albert F.V. van Katwijk, Institute for Perception Research,  
Eindhoven, The Netherlands

The question how the perception of one's own speech may affect the ongoing speech production has been extensively discussed in connection with stuttering, with the effects of delayed auditory feedback and other sidetone monitoring phenomena. There are seemingly conflicting observations in this field: On the one hand stutterers tend to stop stuttering if they are prevented from hearing their own speech, which would suggest that feedback of their own speech is somehow interfering with the production process, whereas on the other hand stuttering tends to occur at the moments of initiation of speech units where auditory feedback is absent.

Most stutterers probably have acquired compensatory or alternative production routines from which it is difficult to disentangle the direct effects of auditory feedback. We therefore made use of subjects who were fluent speakers, in an experiment where we tried to elicit specific feedback effects making use of delayed auditory feedback (DAF).

The main question was: can specific parts of delayed speech be shown to re-enter the ongoing speech production process?

Listening to the performances of 12 subjects repeating a total of 28 three-syllable nonsense words each, there occurred 34 specific misproductions with earlier elements inserted in later parts of the word under production. These misproductions were repetitions of whole syllables and of single vowels. The inserted elements occurred between syllables (repeated syllables) or between C and V. An example of syllable repetitions would be /dadadada/ for /dadada/, and of vowel insertions: /patukui/ for /patuki/. The observed misproductions suggest that the length of the feedback loop determines the probability of a repetition, together with the spots in the articulatory programme where insertions are at all possible.

The main effect of DAF - lengthening - occurred only in second and third syllables. Lengthenings make identifiable impressions on the observer, but are otherwise difficult to interpret in terms of actual processes in the production mechanism.

## THE SELECTION OF PHONETIC TARGETS

Eric Keller, Département de Linguistique,  
Université du Québec à Montréal, CP 8888, Montréal, Qué., Canada

Much recent discussion in phonetics has centered on the problem of co-articulation. This has given rise to a number of theoretical proposals concerning the mechanism which is responsible for anticipatory and perseveratory effects in phonetic production. To expand on these proposals, we have examined anticipatory and perseveratory phonetic effects occurring in aphasic speech, and have evaluated CVC interaction with respect to tongue height by means of large samples of ultra-sound-measured articulations.

From these considerations, the following hypothesis has emerged. The phonetic production mechanism consists of a phonetic target selector and a space coordinate system (cf. MacNeilage, 1970). This production mechanism can perform the selection and implementation of articulations relatively independently of the formulation of the intended utterance, but is constrained by assimilation rules, language-specific segment constraints, co-occurrence rules, and syllabic information. This operation probably occurs in a different time frame from the formulation of the intended utterance. There is therefore a need for a matching operation between the two time frames, which permits correct co-articulatory behaviour in normal speech and provokes anticipatory and perseveratory effects in aphasia.

References

- MacNeilage, P.F. (1970): "Motor control of serial ordering of speech", Psychological Review 77, 3, 182-196.



A COMPLEMENTARY RELATIONSHIP BETWEEN LIP AND JAW MOVEMENTS  
DURING ARTICULATION

Chin-W. Kim, University of Illinois, Urbana, Ill., USA and  
Han Sohn, Yonsei University, Seoul, Korea

It is generally considered that, since the lower lip is attached to the jaw, lip movement is parallel to that of the jaw.

Observations, by means of cineradiography, of timing and distance of the labial and mandibular movements during running speech suggest that, while the two mechanisms behave in coordination, they also behave in a complementary manner in certain cases.

It was noted, for example, during the  $V_1CV_2$  portion of utterances where C is a labial stop, that while the lower lip moved up (for the closure of a labial stop) and down (for the release), the jaw remained stationary during these lip movements but was moving during the labial closure, i.e. while the lip position was stationary. This complementary relationship is shown below:

	$V_1$	C	$V_2$
Lip	Moving	Stable	Moving
Jaw	Stable	Moving	Stable

Two explanations are possible. One is to suggest that while the lips are actively engaged in consonantal articulation, the jaw is rather passive during this time but is active in the positioning of the tongue for the target vowel height. This supports a view (e.g. Lindblom 1967) that there is a closer relationship between vowel height and jaw height than between the former and tongue height.

Another possibility is to attribute the observed phenomenon to mandible coarticulation, i.e., during the labial occlusion the jaw moves, in anticipation, toward the position of the following vowel. Since vowel articulation is in fact facilitated in just this way, one can even argue that mandible coarticulation is an organized principle in speech production.

Further study is needed to determine whether the complementary relationship observed here between the labial and mandibular movements is attributable to this organizing principle of coarticulation or to their differential behavior with respect to consonants and vowels. It is our hope to include in the final report (come August 1979) a detailed comparison of the labial and mandibular movements with that of the tongue so that the question raised may be resolved.

Reference

Lindblom, B. (1967): "Vowel duration and a model of lip-mandible coordination", STL-QPSR 4, 1-29.

TEMPORAL CHARACTERISTICS OF COARTICULATION BETWEEN CONSONANTS AND  
ADJACENT VOWELS - X-RAY MICROBEAM STUDY ON JAPANESE AND ENGLISH -

S. Kiritani, H. Hirose and M. Sawashima, Research Institute of  
Logopedics and Phoniatics, University of Tokyo, Japan

Temporal patterns of coarticulation between consonants and adjacent vowels may vary depending on the phonetic types of the segments. Different languages may also exhibit different temporal characteristics. The present study is an attempt to investigate these problems by observing articulatory movements using the x-ray microbeam method (Kiritani et al, 1975) both on Japanese and American English.

Method

Speech materials studied were  $mV_1CV_2ae$  ( $C=m,t,k,s$   $V=i,e,a,o,u$ ) in Japanese, and  $pVp$  ( $V=$ ten English vowels) and selected CVC words in English. Three or four lead pellets were attached to the tongue and a single pellet was attached to the lower incisor and to the lower lip. Movements of the pellets were tracked by the x-ray microbeam at a rate of 130 frames per second. Pellet positions at selected moments of the consonantal events were sampled and the variations over different vowel contexts were analyzed.

Results and Comments

It was observed that, in Japanese, perturbations of the consonant articulations by the post-consonantal vowels were generally greater than that by the pre-consonantal vowels. The degree of the temporal overlap of the consonant and vowel articulations appears to vary depending on the type of the vowel. Tongue movement for the vowel /i/ showed a greater overlap with consonant articulations than other vowels.

In English, perturbation of the consonants by the pre-consonantal vowels were greater than that by the post-consonantal vowels. The asymmetry between the carryover effect and the anticipatory effect was larger for the so-called tense vowels than for lax vowels.

Effects of prosodic factors such as the stress pattern in English are also being analyzed.

Reference

Kiritani, S., K. Itoh and O.Fujimura (1975): "Tongue-pellet tracking by a computer-controlled x-ray microbeam system". JASA 57, 1516-1520.

PHOTOELECTRIC AND VIDEOFLUOROGRAPHIC REGISTRATION OF VELAR  
HEIGHT: CALIBRATION OF THE VELOGRAPH

Hermann J. Künzel, Institut für Phonetik, University of Kiel,  
Kiel, Federal Republic of Germany

Gaining insight into the velopharyngeal opening-closing mechanism is important for speech scientists, speech therapists and speech pathologists. In order to give these researchers a simple, reliable and at the same time inexpensive tool for investigation the velograph was developed, a photoelectric probe working on the principle of light reflection from the velar surface (Künzel 1977).

So far, the velograph has only been used for the registration of velar timing and relative velar height since the output of the probe in terms of absolute velar height with reference to a baseline had not been calibrated. This procedure is the subject of the present paper. It will be shown that there are high positive correlations between the output of the velograph and velar height gained from simultaneous lateral X-ray video pictures, both for utterances by the same speaker and by different speakers.

Thus, allowing for a certain tolerance interval, real-time registration of velar height may be obtained by using quite a simple instrument. The limitations of the velograph and implications of the technique for future investigations are discussed.

Reference

Künzel, H.J. (1977): "Photoelektrische Untersuchung der Velumhöhe bei Vokalen: erste Anwendungen des Velographen", Phonetica 34, 352-370.

## GENERATING VOCAL TRACT SHAPES IN CONTINUOUS SPEECH

Peter Ladefoged, Mona Lindau and Pat Coady, Phonetics Laboratory, Department of Linguistics, University of California, Los Angeles, California 90024, USA

We will present a film which shows the generation of vocal tract shapes from acoustic data in continuous speech. The display we are trying to generate is roughly equivalent to a traditional midsagittal view of the vocal tract. The shape of the vocal tract is considered to be dependent on seven parameters, each of which may be predicted from acoustic data.

The position of the body of the tongue is defined in terms of two parameters: the amount of raising/lowering of the front part of the tongue, and the amount of raising/lowering of the back part of the tongue (Harshman et al. 1977). These two components can be combined to produce the tongue positions of all vowels and consonants that depend on the position of the body of the tongue.

The position of the tip of the tongue is specified by a third parameter. The jaw and lower teeth are controlled by a fourth parameter. Two further parameters are required to specify the height and width of the lip opening. The position of the velum constitutes a seventh parameter. The values of these physiological parameters are predicted from formant frequencies by a set of equations.

In order to assess the viability of this system, recordings were made of three subjects saying a number of simple phrases. The first three formant frequencies were determined at 10 msec intervals using a computerized LPC formant extraction system and spectrograms.

Given appropriate formant frequencies, plausible sequences of movements of the vocal organs were generated. Since the same set of formant frequencies can correspond to different vocal tract shapes, no claim can be made that these particular movements were used by these particular speakers. But, throughout most of the utterances, vocal tract shapes were generated that could have produced the observed formant frequencies.

Reference

Harshman, R., P. Ladefoged, and L. Goldstein (1977): "Factor analysis of tongue shapes", JASA 62, 693-707.

## THE EPIGLOTTIS AS AN ARTICULATOR

Asher Laufer, Hebrew Language Department, Hebrew University, Jerusalem, Israel, and I.D. Conday, Department of Linguistics, University of Hawaii, Honolulu, HI 96822, USA

We find that the epiglottis functions as an articulator in the production of (1) pharyngeals (2) the vowel [a] (3) whisper. In pharyngeals we find the epiglottis articulates against the posterior pharyngeal wall; the constriction varies from a full closure (pharyngeal stop) in the extreme case of [ʕ] in slow careful speech, through narrow opening (fricative [ħ, ʕ]) in connected speech to fairly open glide [ʕ]. The epiglottis folds toward the pharyngeal wall independently of the tongue root in these consonants. In the vowel [a] the opening is of the same shape as for the pharyngeal consonants, but the opening is substantially larger. The opening allowing the escape of air is between the epiglottis and the pharynx (never between the tongue and the pharynx lateral to the epiglottis). The independence of the epiglottis from the tongue is seen in some cases and not in others for [a]. In whisper the epiglottis is in general more retracted than during normal speech. These observations are based on approximately 100 minutes of videotape made using a fiberoptic positioned in the upper pharynx (of nine subjects), spectrograms, and dissection of cadaver materials.

[This work was supported in part by NIH grant NS9780, the UCLA Phonetics Laboratory, and by Faculty of Humanities of Hebrew University of Jerusalem.]

## THE NEUROMUSCULAR REPRESENTATION OF SPEECH

John Laver, Department of Linguistics, University of Edinburgh, Scotland

Errors in speech which break phonetic realization rules can yield important insights into the nature of the neuromuscular representation of speech.

An experiment<sup>1</sup> is described which provoked vowel-errors of this sort. The random sequencing of two stimuli, and the durations and intervals of their presentation to subjects, were controlled electronically. Stimuli were words of the form P \_ P, containing a stressed vowel of Received Pronunciation, making a list of ten words arranged in 55 different pairs. Each pair was used in a 30-second trial, with the stimulus-duration and inter-stimulus interval both set at .3 sec for the first 15 seconds, then shortened to .2 sec. The task of each of 6 subjects was to pronounce the stimulus-word as accurately as possible immediately on its presentation.

Many vowel blends were produced. Some pairs of vowels were more susceptible to blending than others. An explanation is advanced which ascribes primary responsibility for the execution of a given vowel to a specified muscular system. Vowels blend only when their performance is normally achieved by different muscular systems, the intermediate vowel being the mechanically joint product of both systems acting simultaneously. When two vowels are normally performed by the same muscular system being adjusted to different degrees, then blends don't seem to occur, presumably because individual muscles cannot be given simultaneously contradictory commands.

The general principle of neuromuscular compatibility underlying this argument is clearly also applicable to the study of a number of areas: co-articulatory phenomena, natural classes in phonology, physiologically-motivated sound-change, and physiologically-based constraints on language-acquisition and second-language learning.

---

(1) A fuller account of the experiment reported here will be published in Dechert, H.W. and Raupach, M. (Eds.) (1979) Temporal Variables in Speech: Studies in Honour of Frieda Goldman-Eisler, The Hague/Paris: Mouton.

## A CROSS-LINGUISTIC STUDY OF LIP POSITION IN VOWELS

Wendy Linker, Dept. of Linguistics, Phonetics Laboratory,  
UCLA, Los Angeles, California 90024, USA

For most languages there is a correlation between rounding and backness of vowels, and the degree of lip opening is related to vowel height. For English, lip positions can be predicted from formants following the extraction of only one lip factor (Linker 1978). This factor represented a combination of lip protrusion and vertical lip opening. In languages with both front rounded and unrounded vowels we may hypothesize that different factors are needed to account for the variance in the data. 4 languages with such vowels were investigated (Swedish, Cantonese, French, Finnish) to see how vowel quality interacted with lip position. Two possibilities suggest themselves: (1) these languages conform to a universal relation between lip position and vowel quality, or (2) the relation between lip position and vowel quality is language specific, and the lip positions would differ significantly even if the vowel qualities were identical.

Simultaneous frontal and lateral photographs were taken of 8 male speakers of each of the 4 languages pronouncing words illustrating the vowels of their language. The session was recorded and the first 3 formants of the vowels were measured on spectrograms. The negatives were enlarged and traced. Exact distances for 24 lip measures were calculated. Harshman's PARAFAC procedure, a 3-mode factor analysis, was carried out separately for each language. To compare the languages within the factor space, a version of Canonical Correlation was used, which found a single space for the lip data of all 4 languages. The relation between formants and lip position could then be determined for each language and, since the factor space was identical, meaningful comparisons among the predicted lip spaces could be made by a 3-way Analysis of Variance. Since the lip spaces were predicted from formants, any significant differences are due solely to underlying differences in the use of lip position in these languages.

Reference

Linker, W. (1978): "Lip position and formant frequency in American English vowels", UCLA Working Papers in Phonetics 41, 20-25.



INTER-ARTICULATOR PROGRAMMING IN THE PRODUCTION OF SWEDISH  
OBSTRUENTS

Anders Löfqvist, Department of Phonetics, Lund University, Lund,  
Sweden

Much work within the area of motor control in speech has been devoted to the problem of temporal and spatial coordination of the movements of the various articulators. The production of voiceless obstruents requires a precise temporal control and coordination of several articulatory systems. The tongue, the lips, and the jaw are engaged in the formation of the constriction or occlusion; the soft palate is elevated in order to close the entrance to the nasal cavity and prevent air from escaping that way; the glottis is abducted in order to prevent vibrations of the vocal folds. The present paper reports on some work aimed at elucidating certain aspects of motor control during the production of Swedish obstruents and obstruent clusters.

Registrations comprised a photoglottogram for information on glottal movements, oral egressive air flow and oral air pressure for information on supraglottal articulations, and the signal from a larynx microphone. These registrations were further supplemented with EMG recordings from certain laryngeal muscles.

The results indicate the importance of the temporal coordination of oral release and adduction of the vocal folds for the control of aspiration in voiceless stops. In obstruent clusters the glottis has been found to behave in a manner predictable from the aerodynamic requirements for the production of the respective segments, i.e. the need of an egressive air flow during fricatives and periods of aspiration in stops. In some instances of laryngeal coarticulation the results were not in agreement with those expected on the basis of current theories of motor control in speech. Thus, in some obstruent clusters two successive peaks of glottal opening gestures were found where only one would have been expected. The results will be discussed in relation to current theories of motor control in speech and to laryngeal feature specifications for obstruents. The laryngeal articulation for Swedish obstruents would seem to be best explained as a ballistic opening and closing gesture which is intrinsically tied to certain segments. The temporal relationship between this gesture and the supralaryngeal articulations is important, whereas its size would seem to play a minor role.

## ETUDE DES OCCLUSIVES t/d DU FRANÇAIS PAR L'ELECTROPALATOGRAPHIE

Alain Marchal, Laboratoire de Phonétique, Département de linguistique, Université de Montréal, P. Québec, Canada

L'opposition des occlusives homorganiques a été dans toutes les langues du monde et notamment en français l'objet d'un grand nombre d'études. Nous avons abordé ce problème en utilisant une technique encore peu appliquée à ce sujet, soit l'électropalato-graphie. Notre système composé d'un palais à 64 électrodes et directement relié à un mini-ordinateur sera présenté lors de l'ex-posé de cette communication.

Nous avons examiné les déplacements de la langue au palais lors de la réalisation des occlusives homorganiques /t, d/ et /k, g/ du français dits par 5 locuteurs. Nous avons porté un intérêt tout particulier à l'évolution et à l'étendue (surface) de l'appui de la langue qui fournit un type d'information sur la force arti-culatoire des consonnes. Les conséquences acoustiques des mouve-ments de la langue ont été interprétées à partir d'une analyse sonographique.

Cette étude met en évidence les faits suivants: 1) on observe une grande stabilité des contacts lorsque le barrage de l'occlusive est établie; ce qui manifeste une quasi-immobilité de la langue pendant toute la durée de la tenue. 2) le contexte vocalique, s'il joue un rôle quant au lieu d'articulation, ne modifie pas signi-ficativement l'étendue des contacts de la consonne précédente ou suivante. 3) les occlusives sonores se distinguent par la nature particulière de leur tenue: on peut ainsi constater l'absence d'éner-gie dans le spectre alors que l'occlusion articulaire n'est pas complète. 4) les occlusives sourdes possèdent une longue phase im-plosive en position initiale. 5) les données confirment l'ordre habituellement reconnu des forces articulatoires, soit selon le mode; du plus fort au moins fort - les sourdes puis les sonores et enfin les nasales; selon la position, toutes choses par ailleurs égales, la consonne en position initiale est plus forte qu'en posi-tion intervocalique alors que la position finale se révèle la moins marquée. 6) l'accent final du français exerce une grande influence sur l'articulation des consonnes occlusives.

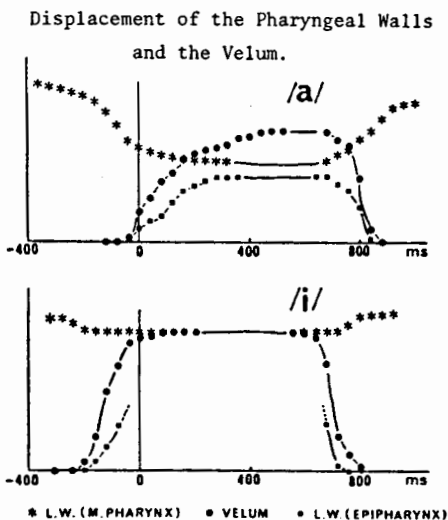
## THE PHARYNGEAL WALL MOVEMENT DURING SPEECH

Seiji Niimi, Research Institute of Logopedics and Phoniatics,  
University of Tokyo, Japan

When we consider that the pharyngeal cavity behaves as a resonator, the movement of the cavity wall, which may act to shape the resonating system, is important and should be investigated as well as the coupling effect of the two resonators: nasal and oral cavities. In this report, the two different levels of the pharyngeal wall (the lateral walls of the epipharynx and the mesopharynx) are studied by means of endoscopy and electromyography.

Results and discussion

The figure shows the displacement patterns of the lateral pharyngeal walls at the two different levels for the Japanese vowels /a/ and /i/. The mesopharyngeal wall moves medially (downward in the figure) to a larger extent for /a/ than for /i/. On the other hand, the medial excursion (upward in the figure) of the epipharyngeal wall is smaller for /a/ than for /i/. This vowel dependent tendency was also observed in the case of CVN syllable strings. It has been reported previously by the author that the movement of the lateral wall of the epipharynx and the vertical movement of the velum are identical in their patterns and caused by the levator veli palatini muscle (Niimi and Bell-Berti 1977).



In this paper I demonstrate that the superior constrictor muscle of the pharynx is responsible for the lateral movement of the wall of the mesopharynx, and this muscle shows the vowel dependent activities.

Reference

Niimi, S.A. and F. Bell-Berti (1977): "An EMG - air pressure - movement study of velopharyngeal closure in speech", 3rd International Congress on Cleft Palate and Related Craniofacial Anomalies 1977, Toronto, Canada.

## INVESTIGATION OF PULMONIC ACTIVITY IN SPEECH

John J. Ohala, Carol J. Riordan, and Haruko Kawasaki, Phonology Laboratory, Department of Linguistics, University of California, Berkeley, California, U.S.A.

Although it has been known since ancient days that the pulmonic system provides the air under pressure required by almost all speech sounds, there is still considerable controversy surrounding the question of whether there is any active short-term pulmonic involvement in the production of specific speech segments (e.g., aspirated stops) or of stressed syllables. Since the only way the pulmonic system can actively contribute to speech production is by varying the volume of the chest cavity, we sought to shed some light on these issues by recording, in three adult male speakers of English, lung volume (as transduced by a whole-body pressure plethysmograph) along with (combined) oral and nasal airflow and the voice signal during a variety of utterance types. Fundamental frequency was extracted and averages formed of all parameters. Figure 1 shows a sample of the averaged data. The lung volume function was characterized by a momentary slowing of the rate of decrement (in comparison to an estimated 'normal' or background rate) during the production of stop closures (upward arrow) and a quickening of the rate of decrement during fricatives, [h], aspirated stop release, and the production of heavily stressed syllables (downward arrow). In all cases but the last, the variation in lung volume could be interpreted as passive reactions to changing lung pressure occasioned by changes in glottal and/or supraglottal impedance. Only heavily stressed syllables were invariably accompanied by active changes in lung volume. (Work supported by the National Science Foundation and the National Institutes of Health.)

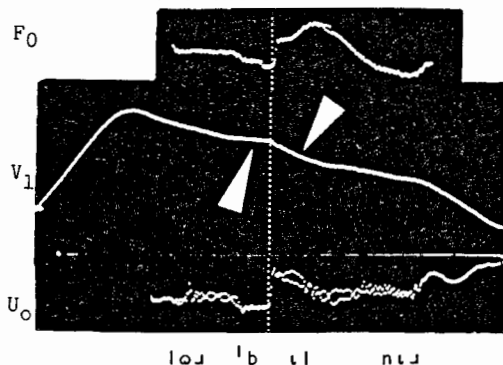


Figure 1. Averaged fundamental frequency ( $F_0$ ), lung volume ( $V_1$ ), and oral airflow ( $U_0$ ) during the utterance 'Lure Bill near', [l o 'b il n l] as produced by an adult male speaker of English. The dotted vertical line marks the synchronization point used to form the averages.

PHENOMENES LIES A L'ENCHAINEMENT DES CONSONNES FRANCAISES  
DANS LA CHAINE PARLEE

Conrad Ouellon, Département de Langues et Linguistique, Faculté des Lettres, Université Laval, Québec, Canada

L'analyse cinéradiologique de consonnes occlusives précédées de voyelles en français permet de préciser l'influence que celles-ci exercent sur ce type de consonnes. Nous examinerons de façon particulière les modifications de lieu d'articulation et de largeur du contact occlusif causées par l'environnement vocalique.

Influence de la voyelle sur le glissement du point de contact des consonnes occlusives

Les consonnes /k/, /g/ et /ŋ/ subissent plus de changements de lieu que les consonnes /t/, /d/ et /n/. L'environnement phonétique peut les expliquer dans chaque cas.

- a) Si la voyelle précédente est antérieure et la voyelle subséquente postérieure, le contact occlusif s'étend ou se déplace vers la partie arrière de la cavité buccale.
- b) Si au contraire la consonne occlusive est précédée d'une voyelle postérieure et suivie d'une voyelle antérieure, le contact occlusif se dirige vers l'avant de la bouche.
- c) Un renforcement d'occlusion provoqué par la présence d'un accent de groupe rythmique sur la syllabe de même que la présence d'une pause après la consonne peuvent entraîner d'autres modifications de lieu de l'occlusion.

Enfin une longue durée favorise les glissements du contact occlusif.

Influence de la voyelle sur la largeur du contact des consonnes occlusives

Les consonnes bilabiales /p/, /b/ et /m/, après une voyelle de faible apertur, tendent à montrer un accolement bilabial plus large.

Un phénomène analogue caractérise les consonnes /t/, /d/ et /n/: à la suite d'une voyelle fermée, la pointe et le prédos de la langue participent souvent à l'occlusion dont la largeur se trouve alors accrue.

Par contre, c'est après une voyelle mi-ouverte ou ouverte que les occlusives /k/ et /g/ montrent un contact plus large. Ce fait s'explique ainsi: les voyelles de grande apertur, dans notre étude, sont postérieures pour la plupart et s'articulent donc dans la même région que /k/ et /g/.

## TEMPORAL COMPENSATION FOR SEGMENTAL TIMING IN ARABIC AND JAPANESE

Robert F. Port, Salman Al-Ani and Shosaku Maeda<sup>1</sup>, Indiana University, Bloomington, Indiana 47401

Temporal compensation may be viewed as the response of the temporal micro-structure of speech (segmental timing) to such macrostructural constraints as constant syllable or word durations. If this hypothesis is correct, we should predict different patterns of temporal compensation in languages with different rhythmic structures. Indeed, study of temporal compensation can be used to illuminate the temporal macrostructure itself. We conducted two similar experiments in Japanese (often cited for regularity of timing) and in Arabic to investigate the compensatory effects of changing the manner and voicing of apical consonants spanning a wide range of constriction durations.

In relevant portions of the Arabic experiment, medial /t,d,r/ in test words were measured in carrier sentences along with preceding and following vowels and the VOT of the initial /k/. Results showed that the voicing change from /t/ to /d/ lengthened the preceding vowel and even VOT, but did not significantly affect either the stop closure itself or the following vowel. The change from /d/ to /r/ resulted in a shorter consonant closure and also in compensatory lengthening of the preceding vowel.

In the comparable experiment on Japanese, two-syllable test words were read in carrier sentences and the durations of all segments in the words measured. Here it was found that /t/ > /d/ > /r/ and all other segments in the test word varied inversely such that total test word durations were the same  $\pm 2\%$ .

These results support the traditional observation of highly regular timing in Japanese but show that the domain of temporal compensation for inherent segmental effects is neither the CV nor VC but rather includes at least two syllables spreading in both directions. Arabic, on the other hand, exhibits far less evidence of temporal compensation and may lack the kind of macrostructure that requires the support of temporal compensation.

---

(1) Also Tenri University, Nara, Japan

## VERTICAL LARYNX MOVEMENT DURING STOP CLOSURE

Carol J. Riordan, Phonology Laboratory, Department of Linguistics,  
University of California, Berkeley, California 94720, USA

An extensive explanatory literature has grown out of the observation that vocal cord vibration is often maintained for longer stop closures than aerodynamic factors apparently allow. Larynx lowering is one of several supraglottal adjustments commonly suggested as a means to prevent equalization of transglottal pressure during oral closure. The empirical support for this hypothesis is primarily a difference in larynx height between voiced and voiceless stops during closure: the larynx tends to be lower for the voiced series, particularly at the moment of release. Earlier reports, however, do not always make explicit that a larynx height difference affects glottal airflow only if it implies an increase in supraglottal volume during the closure interval. This study investigates the change in larynx height during closure and how it relates to oral pressure build-up. Simultaneous larynx height and intraoral pressure records of subjects' productions of intervocalic bilabial stops were measured every 8 msec from 80 msec before consonant closure to 80 msec after release. The results are damaging to the hypothesis that, at least for English, speakers regularly lower the larynx during voiced stops to prolong glottal pulsing. Although previously reported differences in larynx height between voiced and voiceless stops were observed, there were no consistent differences in either the magnitude or frequency of larynx lowering during closure between the two stop categories. Further, the larynx lowered during nasal stops, although nasal airflow presumably maintains transglottal airflow without cavity-enlarging maneuvers. Finally, there was no unique relationship between paired larynx height/intraoral pressure values for the voiced stop as might be predicted.

[Supported by the National Institutes of Health and the National Science Foundation.]

## LARYNGEAL-ORAL COARTICULATION IN GLOTTALIZED ENGLISH PLOSIVES

Peter J. Roach, Department of Linguistics and Phonetics, University of Leeds, Leeds, England

In most accents of British English, glottalization of /p/, /t/ and /k/ is common in contexts other than prevocalic (Roach, 1973). In an experimental study of the articulation of glottalized English plosives laryngeal opening and closing was measured by photo-electric glottograph, and oral closures were detected by electro-palatograph and lip contacts. All instrumental measurements were made and stored synchronously by computer. Laryngeal closure was found to precede oral closure by an average of 8 csec; the time varied according to the magnitude of the articulator movement required to complete the oral closure. There was great variability in the interval between release of the laryngeal closure and of the oral closure. However, there was a very regular relation between release of the laryngeal closure and the onset of a second oral consonant. Considerable timing differences were found between plosive-plus-plosive and plosive-plus-fricative clusters, though the duration of the laryngeal closure was the same in both cases.

Glottalized and aspirated allophones of /p/, /t/, /k/ are in complementary distribution, and it is hypothesized that glottalization has perceptual importance in indicating that the voiceless plosive which it precedes will not be articulated with the extreme glottal opening required for aspiration.

Reference

Roach, P.J. (1973): "Glottalization of English /p/, /t/, /k/ and /tʃ/", J.I.P.A. 3.1, 10-21.



## LARYNGEAL GESTURES FOR VOICELESS SOUNDS DURING SPEECH

Masayuki Sawashima and Hajime Hirose, Research Institute of Logopedics and Phoniatrics, Faculty of Medicine, University of Tokyo

This paper reports some detailed observation of glottal opening and closing gestures for Japanese voiceless consonants and voiceless sound sequences during speech. Fiberscopic view of the larynx was analyzed in correspondence with simultaneously recorded EMG patterns of the laryngeal muscles.

Experimental Procedures

Twelve meaningful Japanese words containing voiceless sounds and sound sequences were pronounced, with a frame sentence, by 2 adult male speakers of the Tokyo dialect. The laryngeal view was filmed using a fiberscope with simultaneous recording of speech and EMG of the laryngeal adductor (INT) and abductor (PCA) muscles. The glottal aperture was measured, frame by frame, on the laryngeal films. A smoothed integrated EMG curve was made for each of the two muscles for each utterance sample.

Results and Comments

The results revealed that the glottal opening varied with different voiceless sounds and sound sequences both in grade and temporal pattern. The results also revealed, at least qualitatively, that the opening and closing of the glottis during speech were controlled by a reciprocal pattern of PCA and INT activity. It should be noted, however, that there was a subject to subject difference in the mode of the laryngeal control using the two muscles. In one subject, the glottal aperture was mainly represented by PCA activity. The time curve of the glottal width in this case could be interpreted as a kind of mechanically smoothed pattern of PCA activity. In the other subject, however, the activity of the INT appeared to actively contribute, in combination with the PCA, to the control of the glottal condition.

(Work supported by Grant-in-Aid for Scientific Research, Ministry of Education No. 349008, No. 239005)

## COMPUTER MODELLING OF ARTICULATOR DYNAMICS

Celia Scully and Edward Allwood, Department of Linguistics and Phonetics, University of Leeds, Leeds LS2 9JT, England

A computer program which models a simple form of the final stages of speech production has been developed. Results obtained with only two articulators, the vocal folds and a single supra-glottal constriction, have shown complex acoustic structures arising from simple articulatory oppositions (Scully, 1975). The model is being extended to generate actual acoustic outputs and to allow interactions between articulators. The response of the vocal tract acoustic tube is being derived by the methods of Husband et al. (1977). The waveform is output in real time via a microprocessor. Each articulator has its own characteristic transition time, constant for large or small distances.

Articulatory gestures are defined by changes in cross-section area of a number of constrictions of the vocal tract. At each 5 msec time sample, the points defining the tongue body, jaw and lips are linked to give a total area function. After a sequence of tongue body shapes has been thus defined, modifications are superimposed; for example, a movement of the tongue tip towards and then away from the palate. Contact appropriate to a plosive will be achieved only if the tongue body position is suitable. Symmetrical or asymmetrical closures and releases may be generated. From individual simple transition functions quite complex total tongue shapes and movements are obtained. The graphs are in agreement with results reported in the literature and with some dynamic palatography data. Diphthong-like sounds have been created from tongue body transitions.

#### References

- Husband, N.M., J.S. Bridle and J.N. Holmes (1977): "A study of vowel nasalization using a computer model of the vocal tract", Proc. Acoust. 9.
- Scully, C. (1975): "A synthesizer study of aerodynamic factors in speech segment durations", in Speech communication, G. Fant (ed.), vol. 2, 227-234, Stockholm: Wiley.

#### Acknowledgments

Supported by the Science Research Council, project number GR/A/19860. Help with modelling the response of the vocal tract from colleagues at the Joint Speech Research Unit and with dynamic palatography from Reading University is gratefully acknowledged.

## ORAL AND NASAL OUTPUTS FOR VOWELS IN NON-NASAL CONTEXTS

Celia Scully and Marion A. Shirt, Department of Linguistics and Phonetics, University of Leeds, Leeds LS2 9JT, England

Several studies have shown that the soft palate is lower for low vowels than for high ones. Vowels with the tongue high in the oral cavity are more susceptible to nasalisation than are those in which the tongue is low, for a given coupling area at the velopharyngeal port, on acoustic grounds, as demonstrated by Fant (1960, 43). But it may also be that the coupling area between nasal and oral tracts is greater for open vowels because the low jaw position drags down the soft palate. Hyde (1968) used a 'nose trumpet' to record separately the acoustic outputs from nose and mouth. He showed that there was a significant nasal output in sounds not requiring a raised velum.

In this study, oral and nasal outputs were obtained using adjacent recording rooms connected by a nose-shaped opening. The speakers were 8 young 'normal' British adults. From the sentences recorded, words containing open and close vowels away from nasal consonants were selected. Separated airflow outputs for these vowels showed many clear cases of aerodynamically non-nasal vowels in these non-nasal contexts. Peak intensity was measured for the oral and nasal acoustic outputs. Two speakers produced the same oral-nasal ratio for both 'open' and 'close' vowels (as judged auditorily), with a mean value of 20 dB. The other 6 speakers gave a smaller oral-nasal ratio of 17 dB for 'open' vowels, with an oral-nasal ratio of 21 dB for 'close' vowels. The results suggest that the soft palate is not severely dragged down, except perhaps for the diphthong /ai/. Some speakers, at least, appear to maintain a constant ratio of oral to nasal acoustic output. The oral output alone sounded 'denasalised'. It seems that the output from the nose may be significant, even in cases where the sounds are apparently transmitted across a raised soft palate.

#### References

- Fant, G. (1960): Acoustic theory of speech production, The Hague: Mouton.
- Hyde, S.R. (1968): "Nose trumpet: apparatus for separating the oral and nasal outputs in speech", Nature 219, 763-765.

METHODE DE SYNCHRONISATION IMAGE-SON POUR L'ETUDE RADIOLOGIQUE  
DES FAITS DE PAROLE - APPLICATION AU FRANCAIS

Péla Simon, André Bothorel, François Wioland et Gilbert Brock,  
Institut de Phonétique, Université des Sciences Humaines,  
22, rue Descartes, 67084 Strasbourg Cedex, France

L'exploitation des films radiologiques est grandement facilitée lorsqu'on dispose d'un enregistrement magnétophonique réalisé simultanément à la prise de vues. Cependant, la synchronisation image-son restant approximative, il n'est pas possible d'établir une correspondance parfaite entre un segment articulatoire et sa réalisation acoustique.

Nous présentons une méthode où cet inconvénient est éliminé: dans le circuit d'enregistrement est intégré un synchronisateur qui, délivrant des impulsions modulées synchrones de chaque image, permet de réaliser une synchronisation graphique entre l'image et le son. L'intérêt de cette technique pour une étude phonétique est grand si l'on considère que la correspondance entre l'image et le son, étant parfaite, a pour avantage de permettre une délimitation rapide du radiofilm, d'établir la superposition entre chaque segment articulatoire visualisé et sa réalisation acoustique.

Nous illustrons l'intérêt que présente cette technique en l'appliquant à l'étude des faits de jointure en français.

Références

- Brock, G. (1977): "Méthode de synchronisation graphique images-son pour l'exploitation des films radiologiques. Présentation de l'appareillage réalisé à l'Institut de Phonétique de Strasbourg". Travaux de l'Institut de Phonétique de Strasbourg 9, 221-232.
- Simon, P., G. Brock et Han Mun-Hi (1977): "Description et utilisation d'un équipement à rayons X pour l'étude de certains aspects articulatoires. Application au coréen" in Modèles articulatoires et phonétiques, R. Carré, R. Descout et M. Wajskop (éd.) 223-242 GALF. Groupe de la Communication Parlée.

METHODE D'ANALYSE DES DONNEES RADIOCINEMATOGRAFHIQUES POUR  
L'ETUDE DES FAITS DE JOINTURE EN FRANCAIS

François Wioland, Laboratoire de Phonétique, Université des  
Sciences Humaines, Strasbourg, France

Cette communication présente une méthode d'exploitation des données radiocinématographiques que nous utilisons pour l'étude d'un aspect des faits de jointure (Lehiste 1965) en français à partir d'un corpus de 65 phrases qui a servi à la réalisation de trois films radiologiques (66 im./sec.). L'utilisation d'un synchronisateur graphique image/son a permis une délimitation très précise des images. La réalisation grandeur réelle de 3500 croquis a diminué sensiblement les inévitables erreurs de reproduction.

A partir d'une image choisie comme référence - organes en position respiratoire - nous avons élaboré une grille de mesure comportant des axes de référence orientés et fixes qui permettent d'établir une quinzaine de mesures pour déterminer la position de la langue, 7 pour la position des lèvres, 4 pour la position du maxillaire, 5 pour la position du voile du palais, ainsi que les coordonnées de l'épiglotte, de l'os hyoïde, du ventricule de Morgagni, de la plaque cricoïdienne et la longueur du canal buccal.

La qualité des films, la précision des croquis et leur parfaite synchronisation avec l'enregistrement justifient à nos yeux le nombre relativement élevé de mesures simultanées qui seul permet par comparaison (Perkell 1969) non seulement de caractériser les indices articulatoires de la jointure en français, mais également d'établir une hiérarchie des différences observées.

L'apparente variété des faits observés jusqu'à présent paraît indiquer l'influence prépondérante, pour un même type de jointure, de la structure phonotactique concernée.

Références

Lehiste, I. (1965): "Juncture", Proc.Phon 5, 172-200.

Perkell, J.S. (1969): Physiology of Speech Production: Results and Implications of a quantitative Cineradiographic Study, Cambridge and London: M.I.T. Press.

## THE PALATALIZATION OF ALVEOLAR FRICATIVES IN AMERICAN ENGLISH

Victor W. Zue, Massachusetts Institute of Technology, Cambridge, MA  
Stefanie Shattuck-Hufnagel, Cornell University, Ithaca, NY

The palatalization of alveolar consonants across word boundaries, as in "got you" (/gat<sup>h</sup>yu/ → [gač<sup>u</sup>]), is a common phenomenon in casual American English. As part of a larger study of this process, we examined the acoustic-phonetic characteristics of the alveolar fricatives /s/ and /z/ in palatalizing contexts. The inquiry was focussed on two issues: (1) in which phonetic contexts can palatalization occur? and (2) how do the acoustic correlates of the resulting palatalized fricative compare to those of the palatals /š/ and /ž/?

Method

The speech material, collected from six speakers, contained many examples of single fricatives, as well as across-word-boundary sequences like /-s##s-/, /-š##š-/, /-š##s-/, /-s##š-/ and /-s##y-/. Measurements, made on both wideband spectrograms and a computer display of waveforms and spectra, included (1) duration, (2) an estimate of spectral concentration at the midpoints of segments, and (3) an estimate of the frequency onset of turbulence noise for midpoint spectra.

Results

On both spectral measurements, the sequence /šs/, as in "tuna-fish sandwich", shows a clear shift from /š/-like values to /s/-like values, while the reverse sequence /sš/, as in "gas shortage" remains approximately constant at values near those for /šš/, /š/ and /s<sup>y</sup>/. This suggests that /sš/ can be palatalized to a single articulatory gesture, while /šs/ requires two discrete gestures, a conclusion which is consistent with spectrographic observations and transcriptions which indicate a single homogeneous fricative for /šš/, and two distinct fricatives for /šs/. Moreover, the duration of /sš/ is shorter than /šs/, as expected if /sš/ merges to a single gesture. A similar pattern of results is found across the voicing variable. Possible explanations of this asymmetry are discussed from the point of view of low-level phonetic rules and articulatory constraints.

## CO-OPERATIVE VOWELS AND COMPETITIVE CONSONANTS?

W. A. Ainsworth, Dept. of Communication & Neuroscience,  
University of Keele, Staffordshire, U.K.

With dichotic presentation Studdert-Kennedy and Shankweiler (1970) found that identification was significantly better for the plosive presented to the right ear, but not for the vowel.

In the present experiments, the formants were split between the ears, in order to discover whether the information combines, as in normal binaural listening, or competes as in the above experiment.

Method

Two sets of nonsense syllables were generated by a synthesis-by-rule system. One consisted of /i, ε, a, ɔ, u, ə/ in an /h-d/ context, and the other of /p, t, k/ combined with each of the above vowels. Listeners identified these syllables in two modes: binaurally, and with F1 + F3 presented to one ear and F2 + F4 presented to the other.

Results

With the vowels, and with /p/ and /t/, no difference was found between the modes of presentation, but with split-formant presentation /k/ was often confused with /p/ or /t/.

Discussion

The results suggest that with split-formant presentation the brief bursts at the start of syllables are analysed independently by the two hemispheres of the brain, the results then compete in order for recognition of the consonant. During recognition of the vowel, however, the information from the two ears is combined.

References

Studdert-Kennedy, M. and D. Shankweiler (1970): "Hemispheric specialization for speech perception", J.A.S.A., 48, 579-594.

## ON THE PERCEPTIBILITY OF MORPHOLOGICAL COUPLINGS IN ENGLISH

Björn Stålhane Andrézen, Institute of Phonetics, University of Bergen, Bergen, Norway

In order to find out how far a number of morphological couplings are auditorily perceptible in English, the following experiment was made:

Groups of sentences were composed, each group consisting of two or three sentences, altogether 34 sentences. Each group was constructed in such a way that the sentences in it contained a stretch of segmental phonemes that was the same in all of them, but with morphological couplings in different places. E.g.: "He was a captain ..." vs. "It was wrapped in ...". The sentences were read on to tape in random order, and then partially deleted, so that of each sentence only the segments that were identical within the group remained. The thus mutilated sentences were played back to a number of listeners of various categories. The listeners had the original texts before them, and they were asked to decide which sentence each fragment had been taken from.

The proportion of correct identifications was higher than pure chance. It seemed to some extent to depend on the relation between the position of the coupling and the consonant(cluster), and on the quality of the consonant(cluster) itself.

#### References

- Bloch, B. (1948): "A set of postulates for phonemic analysis", Lg. 24, 3-46.
- Bloch, B. and G.L. Trager (1941): "Syllabic phonemes", Lg. 17, 223-246.
- Hill, A.A. (1962): "Various kinds of phonemes", Studies in Linguistics 16, 3-10.
- Jones, D. (1967): The phoneme, Cambridge: W. Heffer and Sons Ltd.
- O'Connor, J.D. and O. Tooley (1964): "On the perceptibility of certain word-boundaries", in In honour of Daniel Jones, D. Abercrombie (ed.), 171-176, London: Longmans.



## THE ROLE OF CONTEXT IN VOWEL PERCEPTION

Peter F. Assmann, Department of Linguistics  
University of Alberta, Edmonton, Alberta, Canada

There is considerable evidence that vowel quality is determined largely by the frequency values of the first two formants. However, these values are known to differ between speakers. Other factors have been suggested for the English vowels eg. duration, diphthongization, fundamental frequency and higher formants. Alternatively, contextual information may be involved. Several normalization hypotheses have been proposed. One is that listeners make use of the relationship between formants of different vowels from the same speaker. A second hypothesis states that consonantal or prosodic context provides essential vowel information.

Strange et. al. (1976) emphasize the insufficiency of vowel-internal cues: high error rates are obtained for isolated vowels but not for CVC syllables, in both single-speaker and randomized multi-speaker conditions. Yet Kahn (1977) finds that his subjects make very few errors in the randomized multi-speaker condition. The present study investigates this discrepancy in terms of the following: variability in production, orthographic interference, training and task familiarity and dialect control. When these factors are controlled, listeners make few errors. The increase in errors from CVC's to isolated vowels is attributable to task-related, non-perceptual difficulties.

A second study examines the role of vowel-internal temporal cues. Vowels are artificially shortened by means of a windowing procedure. When temporal cues like duration and diphthongization are removed, errors of identification increase. Confusion errors are reduced when vowels are presented within a block from a single speaker, as compared with a randomized multi-speaker condition. Results are consistent with acoustic measures and lend some support to the relative formant normalization hypothesis. The findings are discussed in terms of "redundant" cues in speech perception.

#### References

- Kahn, D. (1977): "Near-perfect identification of speaker-randomized vowels without consonantal transitions", JASA 62, S101 (A).
- Strange, W., R. Verbrugge, D. Shankweiler, and T. Edman (1976): "Consonant environment specifies vowel identity", JASA 60, 213-221.

## QUELQUES EXPERIENCES SUR LA PERCEPTION DE L'EMPHASE EN ARABE

Belhassen Badreddine, Département de Recherches Linguistiques  
de Paris VII, Paris, France

Le but de cette communication est de démontrer que la corrélation de vélarisation emphatique de l'arabe, interprétée à raison par N.S. Troubetzkoy (1970, 144-45) comme une corrélation de timbre consonantique, est en voie d'être remplacée dans les parlers arabes par une autre plus simple.

Dans le parler d'El-Hamma, Jean Cantineau (1960, 208) remarque que les oppositions d'emphase "souvent ne sont pas constantes; la position principale de pertinence semble être le contact de la voyelle a". Ce qui tendrait à confirmer ce point de vue, c'est le fait suivant: les paires de mots qui attestent l'opposition d'emphase sans que le timbre vocalique a soit suspect d'être l'élément différenciatif sont assez rares dans les dialectes arabes décrits jusqu'à présent.

Pour étudier à fond les restes de l'opposition traditionnelle d'emphase, nous avons songé à recourir à un procédé indirect de vérification en soumettant des listes de paires de mots où l'opposition d'emphase est susceptible d'être pertinente au voisinage de /i/, /u/, /i:/ et /u:/, au sentiment linguistique des sujets parlants.

Les résultats des tests soumis à des sujets tunisiens mettent en évidence le fait que la différence entre un item emphatique et un item non emphatique au contact des voyelles fermées hors situation n'est presque pas perçue par les sujets testés. En revanche, la non-distinction de la différence (confusion) entre les deux phénomènes est très significative. Il est à noter que le pourcentage de confusion est de l'ordre de 70% chez les lettrés et de 90% chez les analphabètes.

#### Références

Cantineau, J. (1960): Etude de linguistique arabe, Paris.

Troubetzkoy, N.S. (1970): Principes de phonologie, Paris: Klincksiek.

## COMPLEX ENCODING IN WORD-FINAL VOICED AND VOICELESS STOPS

William J. Barry, Institut für Phonetik, Universität Kiel,  
Kiel, West Germany

In perceptual studies on the word-final voiced/voiceless distinction, the relationship between vowel and closure duration has been largely ignored. Two experiments report on the effect on the VC dyad of speech rate and position in sentence. Exp. I employs synthetic /bæg(k)/ stimuli of 3 durations (370, 310, 270 ms) simulating "Slow", "Neutral" and "Fast" rates of articulation. The vowel-to-dyad ratios ( $\frac{V}{D}$ ) range from 0.95 - 0.33 in 9 steps per rate. Each stimulus was judged 10 times in 3 separate blocks of 90 stimuli by 20 native speakers of English. Exp. 2 uses 8 computer-manipulated natural utterances: "His bag(ck) seems dirty" and "He's dirtied his bag(ck)" spoken at 2 speeds, "Neutral" and "Fast" with  $\frac{V}{D}$  ratios adjusted in 5 steps to exceed the values of the natural stimulus of the opposing category. Also, the voicing was removed from the natural [bæg] stimulus, producing a voiceless lenis series for each of the 4 tempo-context-combinations. The 4 sets of 15 stimuli were presented, 10 times each, in random order to 13 native speakers.

In Exp. I the  $\frac{V}{D}$  value for the 50 % crossover shifted negligibly from "Neutral" to "Fast" (.70 - .72), indicating an equal perceptual contribution of vowel and closure duration to the combined decoding of phonemic identity and speech rate. The crossover for "Slow", however, occurred at a significantly higher value (.75) than either "Neutral" or "Fast" due to an unchanged closure value from "Slow" to "Neutral". The results of Exp. II confirm those of Exp. I for the sentence-final context, the  $\frac{V}{D}$  values for the "Lenis" and "Fortis" conditions bracketing those of the ambiguous synthetic stimuli: Mid-sentence, however, there was a disproportionately low closure-duration and a correspondingly higher  $\frac{V}{D}$  crossover value for the "Fast" speech rate. This is attributed to the following [s]. A comparison of position in sentence indicates the perceptual importance of sentence-final lengthening, whereby a significant  $\frac{V}{D}$  increase for the "Neutral-Final" combination points to a greater contribution of vowel duration to the juncture signal. All perceptual regularities observed can be linked with corresponding articulatory regularities.

AUDITORY DISCRIMINATION OF RISE AND DECAY TIMES IN TWO  
DUTCH VOWELS

Marcel P.R. van den Broecke, Institute of Phonetics, University of Utrecht, The Netherlands

There are indications that differences in the rise or decay time of the amplitude envelope of vowels may have a distinctive function in some languages as regards their identification. Thus, in French (Malécot, 1975), steep vowel onsets and offsets give rise to glottal stop perception. In Dutch, (Cohen et al. 1963) differences in the decay times of vowels contribute to the perceptual difference between short, half-long and long isolated vowels.

Eight Dutch subjects matched the rise or decay time of a synthesized Dutch /a/ or /ɑ/ with that of a similar reference signal of unknown rise or decay time by means of a blind, 5-turn knob. Rise or decay time of the reference vowel varied between 0 and 100 msec, the invariant slope was fixed at 50 msec. Results show that Weber's Law applies to the responses, i.e. the ratio JND/rise or decay time of the reference signal is constant. The value of this ratio is about 25%.

Accuracy of adjustment, defined as the absolute of the difference between stimulus and response was significantly better in offset position than in onset position. This is due primarily to superior performance in the upper half of the range, 50-100 msec.

The perceptual importance of differences in the decay times in various Dutch vowels may be the cause for this increase in accuracy in the perception of vowel offsets of relatively long duration as compared to offset durations below 50 msec or to onset durations along the entire range used.

#### References

- Cohen, A., I.H. Slis and J. 't Hart (1963): "Perceptual tolerance of isolated Dutch vowels", Phonetica 9, 65-78.
- Malécot, A. (1975): "The glottal stop in French", Phonetica 31, 51-63.

THE EFFECT OF LANGUAGE TYPE ON THE ACUITY OF THE PERCEPTION  
OF DURATION

Arvo Eek, Institute of Language and Literature, Academy of  
Sciences of the Estonian SSR, Tallinn, USSR

The aim of the following experiment is to establish to what extent the just-noticeable difference (JND) of duration changes in case identical stimuli are presented to listeners whose native languages have different word duration patterns.

Stimuli pairs were taken from the vowel /a/, pronounced monotonously and isolatedly, and after removal of its initial transition. All the stimuli begin at one and the same period of the /a/. Eight pair sequences (a+500 ms pause+a) were formed, each of them symmetrical with the reference stimulus. The duration was varied in 2 ms steps and only at the extreme ends of the sequence were the steps increased to 4 ms. The duration of the reference stimulus was increased from 40 to 320 ms in 40 ms steps (in all, 262 stimuli pairs were obtained). The listeners (Estonians - quantity language; Russians - stress language) were asked to mark whether the second /a/ in a pair was longer or shorter than the first /a/. To represent JND, we chose the level of 75% correct responses of the listener's smoothed perception curve, separately for two types of pairs (A - the second stimulus of the pair longer than the first one; B - the second stimulus shorter than the first one).

The JND is larger in the case of reference stimuli with a smaller duration, smaller for medium durations, and shows an increase for larger durations. The  $\Delta T/T$  is largest for smaller durations (15-35%), relatively stable for medium durations (5-7%), and grows towards larger durations, staying, however, within its 40-120 ms region. With Estonians, the difference between the JND's of the A and B pairs is largest in the region of 80-160 ms (with the largest JND in A pairs). Russians have the largest asymmetry in the 160-240 ms region, where the JND is largest in the B pair. It is possible that the a+pause+a pairs are interpreted as disyllabic words (i.e. the pause is identified with a stop consonant). If this is really so, one can ascribe the asymmetry between A and B pairs partly to the different durational patterns of disyllabic words in the respective languages, viz. language specific phonetic structure manifests itself in the discrimination test.

## VARIATIONS IN ATTENTION TO SPEECH: NEW EVIDENCE

John Harris, Institiuid Teangeolaiochta Eireann, Dublin

According to the standard model of speech perception, processing activity is initially geared to the recovery of deep structure from the surface form of sentences. A related claim is that processing activity is concentrated at the ends of clauses as earlier-generated hypotheses are finally resolved.. Evidence for the latter consists of the finding that response latency to a non-linguistic stimulus (a "click") is longer when the click occurs at the end of the first clause than when it occurs at the beginning of the second clause. According to the "on-line interactive" model, in contrast, processing proceeds at all linguistic levels from the first word of the sentence, and the results of earlier processing constrain subsequent processing.

Following this latter model it is claimed that (a) processing activity should not be concentrated at the ends of clauses and (b) processing activity should gradually decrease from the beginning of the sentence as the interpretation of the material becomes more established. Both models lead to the same prediction about differences in latency immediately before and after the clause boundary - the shorter latency is expected at the latter position. In the present study, however, data was collected at all word positions in the sentence and supports both predictions derived from the 'on-line interactive' model. A second question concerned the level(s) of analysis, syntactic/semantic or lexical, which make demands on active attention as measured by click monitoring latency. A comparison of results from the 'click' experiment and results from two earlier experiments (same set of sentences, different linguistic monitoring tasks) provides preliminary answers.

THE PHONETIC FUNCTION OF RISE AND DECAY TIME IN SPEECH SOUNDS,  
A PRELIMINARY INVESTIGATION

Vincent J. van Heuven, Vakgroep Algemene Taalwetenschap, Rijks-  
universiteit Leiden, The Netherlands/ Phonetics Laboratory,  
University of California, Los Angeles, USA

Among the various ways in which speech sounds may differ phonetically or linguistically, such as formant structure, periodicity etc., differences in rise and decay time of the amplitude envelope have received only limited attention in the literature.

In this paper I present a concise survey of the literature, from which two conclusions will be apparent: (a) rise and decay time may indeed contain relevant cues for phonetic/phonemic distinctions, but (b) none of the experiments reviewed safely ruled out all alternative explanations for the effects reported. Moreover, there are hardly any psycho-physical data on the discriminability of rise and decay times, and the results in the only published study on this problem, suggesting increasing sensitivity with longer reference rise/decay times, and categorical perception (Cutting and Rosner, 1974), are counterintuitive.

Before investigating rise and decay time phenomena in a phonetic/linguistic context, however, we felt that more detailed knowledge of JND's of rise and decay time in non-speech stimuli would be in order.

The paper presents the results of our first attempt at establishing these JND's using an adjustment method. Rise and decay times of 1000 Hz sine waves and white noise bursts turned out to have JND's in the order of 25% of the reference signal. Separating out the results for the 4 different signal conditions used (sine/rise, sine/decay, noise/rise, noise/decay) shows that the discrimination curves generally overlap. Performance in the sine/rise condition, however, was slightly better throughout, and a remarkable increase in sensitivity occurs with longer decay times (50-100 msec) of noise bursts. Finally, no traces of categorical perception were found.

Reference

Cutting, J.E. and B.S. Rosner (1974): "Categories and boundaries in speech and music", Perc. Psych. 16, 564-570.

## SOME ACOUSTIC DETERMINANTS OF SYLLABICITY

John T. Hogan, Department of Linguistics,  
University of Alberta, Edmonton, Alberta, Canada

This paper reports a series of experiments on the perception of syllabicity. The first experiment investigates the temporal durations at which 50 percent recognition for syllabicity versus non-syllabicity occurs. The words "stirring", "suing", "bottling", "lightening" and "rhythmic" were recorded by a male Canadian English speaker with the last three words pronounced with a syllabic [l], [ŋ] and [m]. These signals were processed by a PDP-12 computer and the relevant portions of the signal were isolated. The durations of the [ʃ], [u], [l], [ŋ] and [m] in the respective words above were manipulated by a digital gating and editing program to produce signals of four different decreased durations in the syllabic segments. The original plus the four altered signals were presented to fifteen subjects. Crossover boundaries for the five words ranged from 55 to 131 milliseconds. Amplitude increments were made on the shortened durations that occurred in the range where non-syllabicity was perceived. The crossover point was shifted towards a lower duration value at the boundary but no change was observed for the end-point stimuli. An experiment similar to the first was carried out to test whether the loss of tone perception on the syllable occurs within the temporal range of the syllabic/non-syllabic boundary. Finally, temporal summation experiments using the above segments are currently underway to measure the time constant for temporal summation of syllabic segments. Any observed temporal summation with these stimuli may indicate that summation processes are instrumental in the perception of syllabicity.



ADAPTATION IN SYLLABIC CONTEXT: VOWEL CONTINGENT OR  
SPECTRAL SPECIFIC

Peter Howell, Department of Psychology, University College London,  
Gower Street, London, WC1E 6BT, England

A speech adaptation experiment is reported using stimuli which have the same spectral components when cueing a given phoneme before different vowel segments. The stimuli used were consonant-diphthong syllables the diphthongs of which have a rising (/eI/) or falling (/au/) second formant transition. Cooper (1974) has shown that repeated presentation of an alternating sequence of stimuli varying in voicing and in the vowel gives phoneme boundary shifts contingent on the identity of the vowel across the adaptor and test series. One explanation of this result is that vowel contingent feature detectors exist. Another is that adaptation operates on spectral regions.

It is shown that no contingent adaptation effects occur for the stimuli in the present experiment and adaptation occurs in given spectral regions. Further evidence for this conclusion is provided by showing that with one adaptor from a different series, adaptation occurs.

Reference

Cooper, W.E. (1974): "Contingent feature analysis in speech perception", Perc.Psych. 16, 201-204.

## SEGMENTALS AND SUPRASEGMENTALS IN SPEECH PERCEPTION

V.B. Kasevich and E.M. Shabelnikova, University of Leningrad, USSR

Suprasegmentals (intonation, stress, etc.) are generally less directly associated with differentiation of meaning than are segmentals (vowels and consonants). Yet, in Chinese and a number of other languages there exist such apparently indisputable suprasegmentals as tones which are no less important for differentiation of meaning than are vowels or consonants. Our experiments aim at investigating the comparative role played by segmentals and suprasegmentals in Chinese speech perception.

The 1st experiment deals with perception of speech under white-noise masking (signal/noise ratio 0 dB). The intelligibility scores for disyllabic words drawn from arbitrarily chosen sentences show 92.7% recognition for tones and 54.3% for segmentals.

The 2nd experiment studies perception of Chinese speech deprived of its pitch modulations by means of vocoder techniques. Such 'monotonized' sentences presented randomly are found to be 52.6% intelligible.

On the one hand, tones are highly resistant to the effects of white-noise distortion while segmentals are readily confused. This separates the two and affiliates tones with typical suprasegmental behaviour.

On the other hand, the suppression of tones by means of the 'monotonizing' technique is as detrimental to speech recognition as is the 'suppression' of segmentals. This testifies to a functionally common nature of tones and segmentals.

Tones thus appear to be essentially suprasegmental, their function at the same time being non-trivial, sharing much with that of segmentals.

RECOGNITION OF SELECTED PHONETIC CONTRASTS IN THE SPECTRAL AND TEMPORAL DOMAINS BY APHASIC ADULTS

Kurt Kitselman, Veterans Administration Hospital, Martinez, California and University of California, San Francisco

There are conflicting reports regarding the ability of aphasic adults to discriminate phonetic contrasts that are spectral vs. those that are temporal (Carpenter et al., 1973; Blumstein et al., 1977). The present study presented a wider range of phonetic contrasts. Acoustic parameters were manipulated systematically through the use of computer-generated speech stimuli.

Subjects and Methods

Ten aphasic adults and 12 age-matched, neurologically "normal" controls served as subjects. Five 11-item stimulus arrays, each spanning two or more phoneme categories through a succession of equal acoustic changes, were generated. The 5 stimulus categories presented phonetic contrasts that were signaled by (1) 40 msec spectral differences, (2) 25 msec spectral differences, (3) formant transition duration differences, (4) amplitude rise-time differences, and (5) 340 msec spectral differences. Stimulus items were paired within categories at a 2-step level of difference and were presented using an AB discrimination procedure. Response data were pooled across subjects within each group for each of the stimulus categories.

Results and Conclusions

Group discrimination functions differed significantly only for the categories of stimuli that presented brief spectral contrasts. It was concluded that phonetic perceptual disturbances involve disproportionately the brief spectral parameters of speech.

References

- Blumstein, S.E., E. Baker, and H. Goodglass (1977): "Phonological factors in auditory comprehension in aphasia", Neuropsychol 15, 19-30.
- Carpenter, R.L. and D.R. Rutherford (1973): Acoustic cue discrimination in adult aphasia", JSHR 16, 534-544.

## DIMENSIONS IN THE PERCEPTION OF FORTIS AND LENIS PLOSIVES

Klaus J. Kohler, Institut für Phonetik, Universität Kiel,  
Kiel, West Germany

The analysis of the production of fortis and lenis plosives in a great number of languages has shown the importance of the duration ratio vowel/(vowel+closure) for the distinction. Extensive data are presented for German in Kohler et al. (1978).

To complement these results a perception test was carried out in which 29 native German speakers identified a randomised sequence of 220 stimuli from tape as one of the phrases "Diese Gruppe kann ich nicht leid(e)n (leit(e)n)." The stimuli were obtained from the two naturally produced originals by changing the ratios in 6 steps from 0.74 to 0.57 and in 7 steps from 0.55 to 0.79 respectively by computer processing. Similarly 4 steps of consonant voicing were produced in the manipulated leiden-stimulus with the intermediate ratio 0.63, and in the original leiten-stimulus with the ratio 0.55. Each stimulus appeared 10 times in the corpus.

The test results indicate very conclusively that judgment can be reversed simply by changing the ratio to the appropriate ones found in production. Voicing contributes nothing in the case of a clear fortis ratio and only little in an otherwise uncertain area. The psychometric functions for manipulated leiden and leiten are not identical; for the latter it is shifted to higher ratios by 0.08 on average, because the  $F_1$ ,  $F_2$ -transition differences in leiden/leiten are not affected very much by the duration changes applied. Thus a third perceptual dimension determines the identification of fortis and lenis. These dimensions form a hierarchy: duration ratio > formant transition > voicing. The results were duplicated with a second group of 20 subjects.

Reference

Kohler, K.J. and H.J. Künzel (1978): "The temporal organisation of closing-opening movements for sequences of vowels and plosives in German", *Arbeitsberichte Kiel* 10, 117 - 166.

PERCEPTION OF NATURALLY PRODUCED VOWELS: ISOLATED, FROM WORDS,  
AND FROM NORMAL CONVERSATION

Florina J. Koopmans-van Beinum, Institute of Phonetic Sciences,  
University of Amsterdam, The Netherlands

This paper reports on one of a series of studies investigating the influence of various speech conditions or manners of speech on the production and perception of vowels. In a previous study we performed acoustical measurements ( $F_1$ ,  $F_2$ , duration, and fundamental frequency) on the twelve vowels of four Dutch speakers, two male and two female, two trained and two untrained, in eight different speech conditions (non-sustained isolated vowels, vowels in isolated words, stressed and unstressed vowels in a text read aloud, stressed and unstressed vowels in a retold story, and stressed and unstressed vowels in normal, free conversation). Because of the striking 'vowel reduction' in the case of unstressed vowels in normal conversation with reference to isolated vowels and vowels in isolated words, we decided to present these three sets of vowels of each of the four speakers in a listening test to a group of 100 listeners.

Based on their judgments the percentages of correct identifications of the 100 x 216 vowel items for each speaker were:

isolated vowels, resp.	95%, 79%, 88%, 87%
vowels in isolated words, resp.	88%, 79%, 85%, 85%
unstressed vowels in normal conversation, resp.	31%, 29%, 33%, 39%

These results will be compared with results of other studies reported in the literature (Bond 1976, Strange et al. 1976, Kuwahara and Sakai 1972), and further analysis of the errors will be discussed. Besides, we will try to relate these data to the results of the measurements performed on these vowels as reported above.

References

- Bond, Z.S. (1976): "Identification of vowels excerpted from neutral and nasal contexts", JASA 59, 1229-1232.
- Koopmans-van Beinum, F.J. (1976): "Vowel reduction in Dutch", nr. IV, Proceedings of the Institute of Phonetic Sciences, Amsterdam.
- Kuwahara, H. and H. Sakai (1972): "Perception of Vowels and CV-syllables segmented from connected Speech", JAS Japan, 28.
- Strange, W., R.R. Verbrugge, D.P. Shankweiler, and T.R. Edman (1976): "Consonant environment specifies vowel identity", JASA 60, 213-224.

## PERCEPTION ET DECODAGE LINGUISTIQUE: DEUX PROCESSUS DIFFERENTS

Elisabeth Lhote, Laboratoire de Phonétique, Université de Franche-Comté, Besançon, France

Ce travail essaie de dégager deux propriétés importantes de la perception de la parole continue: ce qu'on a l'habitude d'appeler perception en Linguistique recouvre à la fois les mécanismes perceptuels de l'audition et le niveau d'abstraction supérieure qui inclut le décodage linguistique; la structure temporelle des faits émis et celle de leur intégration linguistique chez l'auditeur sont reliées par des lois complexes. Nous avons travaillé exclusivement sur la mélodie intonative de la phrase.

Expériences

Nous avons construit des mélodies synthétiques visant à reproduire les différences tonales qui suffisent en français à opposer des phrases entre elles et soumis deux groupes d'auditeurs différents à des tests:

- a. Nous avons demandé à un groupe d'identifier les patrons linguistiques à partir de ces mélodies (Lhote 1977);
- b. Nous n'avons pas dit au 2e groupe qu'il s'agissait de mélodies de phrases; nous avons demandé aux sujets de dessiner les mélodies (mécanisme perceptuel), puis après avoir pris connaissance des modèles, d'identifier les patrons intonatifs (processus linguistique) (Studdert-Kennedy et Hadding 1973).

Résultats et conclusions

Nous avons dégagé des indices de la perception de l'intonation ayant une fonction prédictive, d'autres ayant une fonction d'intégration, indices qui attestent le décalage qui peut exister entre les faits produits et leur décodage. Ayant observé qu'il y a projection du niveau linguistique sur des attitudes perceptuelles, nous pensons que le niveau linguistique, niveau d'intégration supérieure, impose ses références et ses structures à la perception proprement dite.

Références:

- Lhote, E. (1977): "Quelques problèmes posés par l'élaboration de règles prédictives de l'intonation", Proceedings of the Phonetic Sciences Congress IPS-77, Miami.
- Studdert-Kennedy, M. et K. Hadding (1973): "Auditory and linguistic processes in the perception of intonation contours", L&S 16, 293-313.

## ON THE AMERICAN ENGLISH FLAP

Leigh Lisker, University of Pennsylvania, Phila., Pa. and  
Haskins Laboratories, New Haven, Conn., USA

The flap in American English is phonologically ambiguous and phonetically not well specified. In current parlance it is said to represent either an underlying /t/ or a /d/. For those dialects which distinguish latter from ladder it is generally believed that a difference in the duration of the vowel preceding the flap is the distinctive mark. But it is not true that wherever /t/ + [ɾ], /d/ does likewise. There are varieties of American English where, on the one hand, center includes a flap and sender does not, and where, on the other hand, winter is distinct from winner. In the center-sender pair /t/ is produced with a shorter (= laxer?) occlusion than /d/, - a difference quite the reverse of the situation with the other stops, since /p/ and /k/ are usually stopped for longer intervals than are /b/ and /g/. This center-sender difference makes it hard to understand why linguists ever seriously supposed /ptk/ and /bdg/ of American English to be reliably separated on the basis of a fortis-lenis (= longer-shorter) contrast.

The medial consonant of center is described as a nasalized or nasal flap ([ɾ̃] or [ɾ̃̃]); it contrasts with a nasalized stop [ɳ] in the pair winter-winner. An acoustic analysis of tokens of the two words indicates that the medial closure in winner is longer, and that the signal level during the closure tends to be higher than in winter. Tests in which these closure features were systematically varied did not confirm their perceptual importance for the distinction, even though the durational difference appears to be the main difference between flap and stop articulation, and both duration and signal level are acoustically salient features by which the two words may be distinguished by spectrographic inspection. Instead, other tests showed that listeners' responses are more strongly affected by the presence vs. absence of nasalization in the speech signal at and following release of the constriction.

## SPEECH PERCEPTION IN NOISE

I.M. Lushchikhina, Faculty of Psychology, Leningrad State University, USSR

The role of separate individual audiograms and psycho-physiological peculiarities of the listeners in speech perception in noise via headphones was investigated in three different acoustic conditions: good, average, and bad.

To estimate audiometrical characteristics, methods of tonal and noise audiometry, ear discomfort, and ear stability to sound loads were used. Individual psychological peculiarities of the listeners were estimated according to the Spilberger scale of anxiety (anxiety is considered a characteristic of a person), subjective ideas of listeners about their degree of confidence during perception, and typological properties (strong-weak nervous system).

#### Results

Correlation analysis of results obtained proved a lack of relationship between the listeners' individual features of hearing and their perception in noise.

Anxiety of listeners did not show any connection with the results of perception.

A high negative correlation ( $r = -.78$ ) was found between the property of the nervous system, determined as "weak" and results of perception.

Factor analysis of obtained data proved relative independence of speech perception in noise.

#### References

- Fress, P. and G. Piage (1966): Experimental psychology, Moscow.  
Methodology of investigations of engineering psychology and psychology of labour (1975), Leningrad.



## ON THE IDENTIFICATION OF ARGENTINE SPANISH VOICELESS FRICATIVES

Ana M.B. de Manrique and Maria I. Massone, Laboratorio de Investigaciones Sensoriales, Buenos Aires, Argentina

The present work attempts to examine the perceptual load carried by the frequency position of the most prominent energy-density maximum in the identification of Argentine-Spanish voiceless fricatives. The results are compared with those obtained by Fry (1973) from a group of English-speaking listeners.

Procedure

The test tape consisted of 13 synthetic syllables formed by a fricative voiceless consonant plus a vowel (transitionless), repeated eight times and randomized.<sup>1</sup>

The vowel values were fixed and the fricative portion was obtained by filtering a wide-band noise in order to obtain a set of 13 frequency variable bands ranging from 1.250 to 7.500 Hz.

Two groups of Argentine Spanish-speaking listeners and one of English-speaking listeners were tested under two experimental conditions: free-choice and forced-choice. The latter method was employed in order to allow the comparison between our results and those obtained by Fry.

Results and Discussion

Spanish-speaking listeners identified high and low frequency bands as /f/ and middle ones as /s/. Two noise-bands in between /s/ and low /f/ were sometimes identified as /ʃ/.

These results are not in agreement with those obtained from English-speaking listeners who divided the voiceless continuum in two sections: /s/ for high and /ʃ/ for low frequency values.

Both English and Spanish-speaking listeners' responses were only slightly influenced by the forced-choice condition. Thus, the difference between the two sets of data cannot be accounted for by the method employed and may probably be attributed to a different use of the acoustic properties due to the peculiarities of each linguistic system.

Reference

Fry, D.B. (1973): "Acoustic cues in the speech of the hearing and the deaf", Proc. Royal Soc. of Medicine, 66, 959-969.

---

(1) The authors wish to thank Dr. D.B. Fry for his advice and for providing them with the test tape.

## PERCEPTUAL CENTRES (P-CENTRES)

Stephen M. Marcus, Instituut voor Perceptie Onderzoek (I.P.O.), Eindhoven, Nederland

The generation of perceptually regular sequences from a set of naturally spoken digits stored on a computer poses some fundamental problems in the timing of speech sounds (Morton et al., 1976). It is immediately clear that perceptual regularity does not correspond to regularity of acoustic onsets. In order to investigate what is regular in a "regular" sequence, the PERCEPTUAL CENTRE (P-centre) of a sound is defined as its psychological moment of occurrence. "Regularity" is then, by definition, regularity of P-centres.

It is hypothesized that P-centres are determined only by the acoustic nature of each stimulus, invariant of the context provided by adjacent stimuli. This hypothesis is tested and a paradigm described for the determination of P-centre locations of isolated speech stimuli relative to one another.

The relationship is considered between the results of these experiments and those of Rapp (1971) and Allen (1972). It is concluded that a large component of the variance in their tasks involves individual differences in temporal coordination of speech and non-speech or motor tasks; these differences were absent in this paradigm involving relative timing of speech sounds. Rapp's model of P-centre location is evaluated with data from this paradigm. It is found that although she employs the most important parameter, that of consonant duration preceding the nuclear vowel, vowel and final consonant duration must also be considered as important secondary parameters. Experiments investigating P-centre shifts produced by selective modifications of digitized stimulus waveforms (Marcus, 1976) also show that P-centre location is principally a function of stimulus duration and not of stimulus energy.

References

- Allen, G.D. (1972): "The location of rhythmic stress beats in English. Parts I & II", L&S 15, 72-100; 179-195
- Marcus, S.M. (1976): Perceptual Centres, unpublished PhD thesis, Cambridge University, Cambridge, England
- Morton, J., S.M. Marcus and C.R. Frankish (1976): "Perceptual Centers (P-centers)", Psych.Rev. 83, 405-408
- Rapp, K. (1971): "A study of syllable timing", Speech Transmission Laboratory, Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm.

## PERCEPTION OF STOP CONSONANTS BEFORE LOW UNROUNDED VOWELS

Ignatius G. Mattingly, Haskins Laboratories, New Haven, Connecticut, and University of Connecticut, Storrs, Connecticut, USA, and Andrea Leavitt, Haskins Laboratories, New Haven, Connecticut, USA, and Wellesley College, Wellesley, Massachusetts, USA

Previous experiments in the perception of stop-vowel syllables have sampled the entire vowel space rather coarsely (e.g. Delattre et al., 1955; Harris et al., 1958; Hoffman, 1958; Liberman et al., 1954). The present experiment looks more closely at the perception of stops with four low unrounded vowels differing only in F2 frequency and heard as more or less backed variants of [a].

Labelling Tests

For each vowel, two labelling tests were prepared from synthesized stimuli. The onset of the F3 transition was varied in seven 200 Hz steps centering on the F3 steady state value and the onset of F2 was varied in five 100 Hz steps centering on previously obtained estimates of the [b-d] and [d-g] crossover points for F2 with a straight F3 transition. The tests were given to 12 subjects.

Results

The pattern of crossover values obtained reflects the interaction of the F2 and F3 transition cues and the sharp difference in the velar locus before front and back variants of [a].

References

- Delattre, P.C., A.M. Liberman, and F.S. Cooper (1955): "Acoustic loci and transitional cues for consonants", JASA 27, 769-773.
- Harris, K.S., H.S. Hoffman, A.M. Liberman, P.C. Delattre, and F.S. Cooper (1958): "Effect of third-formant transitions on the perception of the voiced stop consonants", JASA 30, 122-126.
- Hoffman, H.S. (1958): "Study of some cues in the perception of the voiced stop consonants", JASA 30, 1035-1041.
- Liberman, A.M., P.C. Delattre, F.S. Cooper, and L.J. Gerstman (1954): "The role of consonant-vowel transitions in the perception of the stop and nasal consonants", Psychol. Monogr. 68, 8:1-13.

[Support from the National Institutes of Health and the Veterans Administration is gratefully acknowledged.]

## ALLOPHONIC AND PROSODIC CUES FOR PARSING SPEECH

Lloyd H. Nakatani, Bell Laboratories, Murray Hill, N. J., U. S. A.

A theory of speech perception must explain how listeners hear discrete words in a continuous acoustic signal. We show that listeners hear words by dividing and combining stretches of the speech stream -- that is, by parsing speech -- into short word-sized portions which are likely to be actual English words. Parsing is done perceptually from allophonic and prosodic cues, not inferentially from syntactic and semantic knowledge. In this view, speech perception goes from continuous speech to discrete words from the bottom up, not from the top down.

Speech is parsed with the aid of allophonic and prosodic variations which function as either fission or fusion cues. Fission cues indicate portions of the speech which are divided by a word boundary. Examples of fission cues are (1) allophonic variations such as aspiration of word-initial voiceless stops, and glottalized onset of word-initial stressed vowels; and (2) prosodic stress and rhythm cues such as consecutive primary stressed syllables which must perforce belong to different words, and a long stressed syllable which is probably a monosyllabic word or phrase-final syllable and therefore must be followed by a word boundary.

Fusion cues, by contrast, cause portions of the speech to fuse perceptually so that a word boundary cannot divide the portions. Examples of fusion cues are (1) allophonic variations such as the syllabic nasal in "maiden" where the /d/ and /n/ are fused, and (2) prosodic stress and rhythm cues such as an unstressed syllable (other than a function word) which must be part of a polysyllabic word, and a short stressed syllable which is probably a non-final syllable of a polysyllabic word formed by fusion with a following unstressed syllable.

Our experiments show that fission and fusion cues are important for parsing speech. But they are not enough. Listeners probably also hear function words and affixes, and use their knowledge of where these sounds occur in English to parse speech. Experiments are planned to see if function words and affixes are used in parsing.

## FORMANT FREQUENCY VARIATION AND VOWEL QUALITY

T.M. Nearey, University of Alberta, Edmonton, Canada T6G 2H1

Two sources of within-phoneme variation have been of major interest to experimental phonetics: 1) context-dependent and 2) speaker-dependent. The relative importance of these sources of variation is examined in the light of natural data and synthetic speech experiments. It is argued that speaker variation is both greater in magnitude and more systematic than context variation. The phonetic import of physical variation must be carefully considered in evaluating this question.

Mermelstein (1978) provides evidence that much of the contextual variation observed thus far is below threshold, and hence perceptually irrelevant. Lindblom's undershoot model for formant variation has been seriously weakened by recent results reported by Gay (1978). Although limited evidence exists for a perceptual mechanism that could compensate for some contextual variation, (Lindblom and Studdert-Kennedy 1967), the degree of complementarity between natural context variation and perception is not clear.

The magnitude of speaker variation is several times larger than that of context variation. Nearey (1977) provides evidence for a detailed complementarity between natural speaker variation and perception in synthetic stimuli. A "constant ratio hypothesis" (CRH) is shown to provide an excellent fit to natural data. Furthermore, an important perceptual implication of CRH is supported in a synthetic vowel experiment: the change in the formant frequencies of a single context vowel is sufficient to produce a global monotonic shift in categorization boundaries of a vowel continuum that covers F1-F2 space.

#### References

- Gay, T. (1978): "Effect of speaking rate on vowel formant movements", *JASA* 63, 223-230.
- Lindblom, B. and M. Studdert-Kennedy (1967): "On the role of formant frequency transitions in vowel recognition", *JASA* 42, 830-843.
- Mermelstein, P. (1978): "Difference limens for formant frequencies of steady-state and consonant-bounded vowels", *JASA* 63, 572-580.
- Nearey, T. (1977). Phonetic feature systems for vowels. Dissertation, University of Connecticut.

## SPECTRAL AND PERCEPTUAL ASPECTS OF VOWEL COARTICULATION

Louis C.W. Pols <sup>a)</sup> and M.E.H. Schouten, Institute for Perception TNO, Soesterberg, the Netherlands. <sup>a)</sup> presently at Speech Communications Research Laboratory, Inc., 800A Miramonte Drive, Santa Barbara, California 93109

Formant transitions, or more general acoustic characteristics, of vowel transitions in CV- and VC-type syllables are known to carry information about the preceding, or following, consonant as well as about the vowel itself. Although for instance Haskins' locus theory and Lindblom's model give some way of describing these phenomena, based on experiments with synthetic speech, a full description of what is actually occurring in real speech is far from being available.

A study to come up with some of these data should include both acoustic measurements on actual speech, and a perceptual evaluation of the significance of its dynamic characteristics.

Detailed spectral data for a subset of Dutch CV- and VC-transitions are now available both for isolated words and for words in a read-aloud story (Schouten and Pols, 1979). The CV- and VC-transition patterns were found to be quite consistent over speakers and conditions. Perceptual experiments have been conducted to specify the extent to which vowel transitions contribute to the identification of preceding or following plosives in Dutch CVt or tVC words. Large differences were found between initial voiced and unvoiced plosives (Pols and Schouten, 1979).

These experiments will be replicated for a full set of American English plosives. Experiments are also planned to extend these perceptual studies to all consonants. Another laborious but interesting extension is to use running speech, or to isolate stimuli from running speech.

This information will tell us more about the contribution of dynamic speech characteristics to speech perception, and will also contribute to improve automatic speech recognition procedures.

#### References

- Schouten, M.E.H. and Pols, L.C.W. (1979): "Vowel segments in consonantal contexts: a spectral study of coarticulation-Part I", JPh forthcoming.
- Schouten, M.E.H. and Pols, L.C.W. (1979): "CV- and VC-transitions: a spectral study of coarticulation-Part II", JPh forthcoming.
- Pols, L.C.W. and Schouten, M.E.H. (1979): "Identification of deleted consonants", JASA forthcoming.

## PSYCHOAKUSTISCHE FUSION UND DICHOTISCHE ADAPTATION

Bernd Pompino, Institut für Phonetik und Sprachliche Kommunikation der Universität München, Bundesrepublik Deutschland

Die Technik der selektiven Adaptation hat sich als starkes Instrument zur genaueren Analyse der Teilprozesse bei der Sprachwahrnehmung erwiesen. Die vorgestellten Experimente dienen der Klärung der Frage, ob sie nicht auch zur genaueren Erforschung der Hemisphärenunterschiede im auditorischen Bereich verwendbar ist.

Bisher konnten mit dieser Technik keine Hemisphärenunterschiede festgestellt werden. Im bisher einzigen Experiment zur dichotischen Adaptation konnte Ades (1974) aber zeigen, dass sowohl Mechanismen, die über Input von nur einem, wie auch solche, die über einen Input von beiden Ohren verfügen, adaptierbar sind. Der zentrale Effekt konnte durch die Wirkung der spektralen Fusion gezeigt werden. Neben der spektralen Fusion trat in diesem Experiment auch die psychoakustische Fusion auf, die aber bei den Adaptoren /bæ/ vs. /dæ/ zu einem nicht eindeutigen Perzept führen, so dass das Fehlen eines zentralen Adaptationseffekts hier nicht verwunderlich ist.

In unseren Experimenten verwendeten wir daher die Adaptoren /ba/ vs. /ga/ - bzw. deren chirps und bleats -, die die psychoakustische Fusion zu /da/ zur Folge haben. Bei Adaptation mit den vollständigen Silben zeigte sich eine Adaptation an /da/ bei der Adaptorausrichtung /ba/<sub>R</sub> vs. /ga/<sub>L</sub>. Da sich in einem weiteren Experiment /ga/ als stärker gewichtet herausstellte, kann dies als Aufhebung des Effekts der Fusion durch den stärker gewichteten Stimulus am rechten Ohr interpretiert werden. Bei den chirps ergab sich ebenfalls eine Adaptation an /da/, allerdings bei umgekehrter Adaptorausrichtung, wohingegen die bleats in einer unterschiedlichen Adaptation beider Ohren resultierten.

Die Schlussfolgerung aus diesen Ergebnissen ist, dass unter anderem auditive Faktoren die Art der Verarbeitung durch das Nervensystem und die Lateralisierung der auftretenden Prozesse bestimmen.

Literatur

Ades, A.E. (1974): "Bilateral component in speech perception?", JASA 56, 610-616.

## TWO PARAMETERS IN THE PERCEPTION OF SERBO-CROATIAN WORD TONE

Edward T. Purcell, Dept. of Slavic Languages, Dept. of Linguistics,  
University of Southern California, Los Angeles, Calif. 90007, USA

Previous reports by this author have described the realizations of Serbo-Croatian word tones in differing sentence environments (Purcell 1972, 1973). It was found that several patterns of fundamental frequency differences regularly occurred in the accented and first post-accentual vowel, which seemed to differentiate rising and falling tones. It was also reported that differences in the location of the fundamental frequency peak within the accented vowel were observed, which also seemed to differentiate rising and falling tones. In another paper, perceptual data were presented indicating that natives can use such differences in the location of the pitch peak within the accented vowel to discriminate rising and falling tones (Purcell 1976). In the present paper we will present perceptual data comparing two parameters:

1. the location of the pitch peak within the accented vowel and
2. the relationship between the first and last fundamental frequency value within the accented vowel.

Three gradations of peak location were combined with five gradations of start/end ratio in synthetic stimuli. Native listeners' responses were subjected to multiple regression to assess the relative contribution of each parameter to the perception of Serbo-Croatian word tone.

References

- Purcell, Edward T. (1972): "The acoustic differentiation of Serbo-Croatian word-tones in statement environments", Proc.Phon. 7.
- Purcell, Edward T. (1973): The realizations of Serbo-Croatian accents in sentence environments, Hamburger phonetische Beiträge 8, Hamburg: Buske.
- Purcell, Edward T. (1976): "Pitch peak location and the perception of Serbo-Croatian word tone", JPh 4, 265-270.



BIDIRECTIONAL CONTEXT EFFECTS IN PERCEPTION OF SYNTHETIC  
FRICATIVE-(STOP-)VOWEL STIMULI

Bruno H. Repp and Virginia A. Mann, Haskins Laboratories,  
270 Crown Street, New Haven, Connecticut 06510, U. S. A.

In this paper, we describe two examples of context dependency in speech perception--one retroactive, the other proactive--and report a series of experiments conducted to delimit the conditions necessary for their occurrence. The retroactive effect is observed when stimuli from a synthetic /ʃ/-/s/ continuum are followed by different vowels: Listeners report /s/ more often when the vowel is rounded than when it is not (Kunisaki and Fujisaki, 1977). We have replicated this effect using /a/ and /u/ as vowel contexts. We find that the magnitude of the retroactive effect changes little as fricative noise duration is extended, but that it is substantially reduced when silent intervals of varying sizes are introduced between noise and periodic portions. The proactive effect occurs when stimuli from a synthetic /da-/ga/ continuum are preceded by a fricative: Listeners give more velar stop responses following /s/ than following /ʃ/. This effect is remarkably persistent, although its magnitude does decrease with increased temporal separation of noise and periodic portions and with presence of a syllable boundary between fricative and stop.

In both cases, there are certain parallels between our perceptual results and coarticulatory effects in speech production. The retroactive effect corresponds to the effect of anticipatory lip rounding on the spectrum of fricatives preceding rounded vowels (Kunisaki and Fujisaki, 1977), and we have obtained some evidence for a forward shift in place of articulation for stops following /s/, which is consistent with the proactive effect in perception. Thus, speech perception appears to be guided by an implicit knowledge of articulatory dynamics.

References

- Kunisaki, O., and Fujisaki, H. (1977): "On the influence of context upon perception of voiceless fricative consonants", Annual Bulletin of the Research Institute of Logopedics and Phoniatrics (University of Tokyo) No. 11, 85-91.

## LES CONFIGURATIONS ET L'INTERACTION DES PENTES DE Fo ET DE I

M. Rossi, Institut de Phonétique, Aix-en-Provence, Laboratoire Associé au C.N.R.S., n° 261.

Nous nous proposons d'étudier la perception des glissements d'intensité (GT.I) dans la parole, leur action sur les glissandos de fréquence (Go.Fo) et leur mode de perception.

Résultats

Nous avons expérimenté l'influence des GT.I positifs et négatifs sur des tons mélodiques statiques, montants et descendants(1). Un GT.I positif ou négatif associé à un ton statique est perçu comme un glissando de même sens qu'un GT.I et supérieur au seuil. Mais un GT.I négatif associé à un Go.Fo montant ou descendant diminue la perception du ton mélodique, tandis qu'un GT.I positif, dans les deux cas, favorise la perception et affine le seuil. Un GT.I positif a un effet sur la configuration du ton : associé à un Go.Fo montant, le ton est creusé ou concave, et avec un Go.Fo négatif, le ton est convexe. Nous proposons un modèle fondé sur la concordance temporelle des points de hauteur et de phonie.

Dans une nouvelle expérience nous testons la perception de variations de GT.I positifs de 0, 4, 8, 12 et 16 dB. Nous mettons en évidence un double effet de GT.I : a) sur la hauteur, b) et sur l'inflexion du ton. On prédit, grâce à la forme de la fonction psychométrique, la stratégie des sujets. 3 expériences complémentaires confirment l'effet de GT.I sur l'inflexion du ton. Il résulte de ces expériences que le ton creusé provoqué par GT.I positif a une forme qui s'apparente à un palier mélodique suivi d'un glissando, mais sans se confondre avec ce dernier ; il est perçu comme une forme mélodique spécifique imposée par certaines contraintes.

Conclusion

On examine les implications des résultats obtenus dans l'étude prosodique, en particulier pour l'interprétation des paramètres des intonations déclarative et interrogative. On conclut sur le caractère pluriparamétrique de l'intonation et sur la nécessité d'une conversion perceptuelle des données objectives.

---

(1) Rossi, M. (1978): "Interaction between intensity glides and frequency glissandos", L&S (à paraître).

VOICING FEATURES IN THE PERCEPTION AND PRODUCTION OF STOP  
CONSONANTS BY JAPANESE SPEAKERS

Katsumasa Shimizu, Department of Languages, Nagoya Gakuin  
University, Seto city, Aichi-ken, Japan

The present study is concerned with the identification of voiced and voiceless consonants by Japanese speakers using synthetic speech sounds varying along a continuous VOT-scale, and also with the articulatory effects in speech perception.

Subjects

One of the major problems in speech perception is to examine how articulatory and auditory mechanisms are linked to each other. Some experiments have been reported on the articulatory effects in speech perception, but most of them are mainly concerned with place features, not with voicing features. It is suggested that there exists a common mechanism for perception and production of speech sounds (Cooper et al., 1975). In the present study, adaptation effects have been examined in repetitive listening and articulation of six syllables /ba, pa, da, ta, ga, ka/. Repetitive articulation of /ba, pa, ka/ caused a shift of phonetic boundaries in the predicted directions, but there are some differences in the strength of the articulatory effects; that is, the feature detector for voiceless is more sensitive than that for voiced, and the labial detectors are more sensitive than other place features. This may indicate that the detectors for each feature do not necessarily function at a mediating level for perception and production and that there is some separate processing of some feature detectors from articulation.

Conclusion

We are now working on the problem how neural commands in articulation affect the processing of auditory and linguistic information and hope to be able to present more concrete results at the 9th International Congress of Phonetic Sciences.

Reference

Cooper, W.E., S.E. Blumstein and G. Nigro (1975): "Articulatory effects on speech perception: a preliminary report", JPh 3, 87-98.

## PERCEPTION OF VOWEL FORMANT TRANSITIONS

Iman H. Slis, Instituut voor Fonetiek, Katholieke Universiteit, Nijmegen, The Netherlands

In the majority of speech synthesizers the filter parameters are changed stepwise during formant transitions. Demonstrations with vocoders show that continuous speech which is synthesized this way can be of excellent quality. It is not known, however, whether a steplike approximation will also suffice for shorter stimuli. Therefore, we started a series of experiments concerning the detectability of vowel formant transitions within one F1 period.

In a pilot experiment, a formant filter was excited with one pulse. The filter parameters in the reference stimulus were kept constant. In the synthesis of the test stimuli the coefficients of the formant filter were linearly interpolated on a per-sample basis; bandwidth was kept constant. Only rising formant transitions were synthesized. The starting points of the test stimuli were arranged around the frequency of the reference stimulus. This experiment was repeated with four filters in series. These experiments were then extended using stimuli of more than one excitation pulse.

We hypothesized that the auditory impression caused by the test stimuli would be dominated by the duration of the first cycle, the amplitude of which is much greater than that of the remaining ones. Therefore we expected to find almost no perceptual difference between reference and test stimuli which have a first cycle of approximately equal duration. However, the first results show that six subjects were capable of hearing substantial differences between reference and all test stimuli; and that the smallest difference was found when the duration of the third cycle of the test stimulus was equal to the cycle duration of the reference stimulus.

## PERCEPTION DE CRESCENDOS D'INTENSITE EN FIN DE PHRASE

Christel Sorin, CNET, Lannion, France

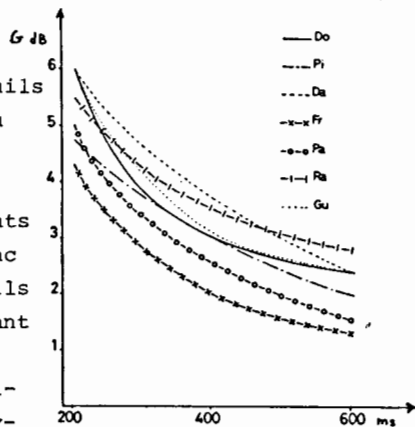
Il est généralement admis que l'étude des faits prosodiques doit se faire au niveau des trois paramètres: fréquence fondamentale ( $F_0$ ), durée et intensité. Nous nous sommes attachés dans l'expérience décrite ici, à déterminer quelle était la précision de la perception des variations d'intensité en fin de phrase, poursuivant ainsi les travaux de Rossi sur les voyelles.

Méthode et expérimentation

Nous avons étudié la discrimination d'un crescendo d'intensité appliqué sur les 200, 400 et 600 dernières ms de phrases naturelles. Six phrases, prononcées chacune par 2 locuteurs (1H et 1F) formaient le corpus. Les phrases étaient présentées par paire et la mesure de discrimination effectuée par la méthode des limites sur 7 sujets. A titre de comparaison, la même mesure a été reproduite sur des signaux de bruit blanc stationnaire, puis sur des signaux de bruit blanc modulés en amplitude par l'enveloppe de chacune des phrases testées.

Résultats et interprétation

On a porté sur la figure les seuils obtenus (en terme de valeur finale du gain  $G$ ) par chaque sujet sur les 12 phrases, en fonction de la durée du crescendo. Contrairement aux résultats obtenus sur les signaux de bruit blanc stationnaire, l'évolution de ces seuils suit une courbe décroissante atteignant vers 500 ms la valeur du seuil différentiel observé sur la parole continue (1 à 2 dB). Diverses mesures physiques effectuées sur le signal pour rendre compte des résultats



subjectifs observés laissent supposer que l'oreille intègre la partie finale du signal à partir de la voyelle la plus intense située à plus de 200 ms de la fin. C'est ensuite sur les valeurs de cette puissance "moyenne" que porte la comparaison. Le rôle éventuel d'un décodage phonétique dans ce traitement sera discuté en comparant les résultats obtenus sur les phrases et sur les signaux de bruit "à enveloppe de parole".

Référence

Rossi, M. (1978): "The perception of non-repetitive intensity glides on vowels", JPh 6, 9-18.

## SCALING OF CERTAIN SELECTED DISTINCTIVE FEATURES IN ENGLISH

James Monroe Stewart and Carol M. Barach, Department of Communication, Tennessee State University, Nashville, Tennessee 37203, USA

The specific purpose of this study was to determine whether or not a hierarchical structure exists within the phonological domain of distinctive features. The secondary purpose was to determine whether the Chomsky and Halle (1968) Distinctive Feature System is relevant to and descriptive of the perceptual domain of the adult listeners in a speech processing mode. At a micro-level of speech perception, the goal of the study was to identify and describe some of the underlying strategies of distinctive feature utilization associated with subjects' perceptual judgements of certain selected speech stimuli.

#### Subjects

The evaluation of the subjects' responses was obtained through combining the experimental tasks of absolute judgment and magnitude estimation of the minimally distinct members of the stimulus sets. A stimulus set consisted of a referent CV-nonsense syllable followed by three target CV-nonsense syllables. The subjects were to order the relative similarity of each of the three target syllables in relation to the referent. The CV syllable was utilized in order to maximize acoustic and minimize linguistic effects.

#### Conclusion

The study supports a hierarchical ordering of the saliency and a perceptual organization with distinctive features. One may also conclude that the Chomsky and Halle (1968) Distinctive Feature System appears to have some relevance and descriptivity of at least some phonemes in English.

#### Reference

Chomsky, N. and M. Halle (1968): The Sound Pattern of English, New York: Harper and Row.

## THREE SOURCES OF INFORMATION IN VOWEL IDENTIFICATION

Winifred Strange and James J. Jenkins, Psychology Department,  
University of Minnesota, Minneapolis, Minnesota, USA

Three studies investigated the sources of information used by listeners to identify vowels spoken in syllabic contexts. Traditional theory holds that target formant frequencies are most important in vowel identification. Recent research suggests, however, that dynamic information plays an important role in determining accurate identification.

Stimuli

Native English speakers recorded b-vowel-b syllables for 9 or 10 vowels. These syllables were electronically processed in various ways to (1) delete the formant transitions, (2) delete the syllable centers, leaving only the initial and final transitions, and (3) distort or eliminate the differential duration information. Separate identification tests were prepared for each condition.

Subjects

Independent groups of naive listeners (college students) attempted to identify the vowels.

Results

Errors were scored if the listener reported other than the intended vowel. Error patterns in all three experiments were highly similar. Unmodified syllables, of course, had the lowest error rate, but syllables from which the center had been deleted were almost as good. Identification of the syllable centers without transitions was somewhat poorer. When these centers were given constant duration, identification was extremely poor. Changing the duration of silence in the syllables which had centers deleted produced an intermediate level of errors.

Conclusion

Formant transitions and durational information are important sources determining accurate vowel identification. Formant center frequencies alone, stripped of dynamic information, are relatively poor sources of identification information.

## ZUR BELEGUNG EINES HIERARCHISCHEN SPRACHPERZEPTIONSMODELLS

W. Tscheschner, Technische Universität Dresden

In den Proceedings des Speech Communication Seminar Stockholm 1974 [1] wurde als Ergebnis psychophysikalischer Experimente und rechentechnischer Simulationen sprachverarbeitender Automaten ein Sprachperzeptionsmodell vorgestellt. Hierbei erfolgt die sequentielle Verarbeitung eines Sprachsignals auf der Basis einer Merkmalabbildung, einer Eigenschaftsdiskrimination und einer logisch orientierten Lautentscheidung über einem physisch bedingten Zeitregime.

Über selektive psychoakustische Perzeptionsuntersuchungen können Modellkomponenten des dynamischen Reaktionssystems untersucht werden. Am Beispiel der subjektiven Vokalerkennung wird erläutert, wie das Merkmal Tonheitslage eines dominant empfundenen Lautheitsmaximums die Zuordnung hinten artikulierter Vokale [u:, o:, a:], unter einschränkenden Bedingungen, vollständig zu beschreiben vermag. Dabei muss das Ergebnis mit bisher statistisch gesicherten Einsichten verträglich sein.

"Volumen" oder auch "Öffnungsgrade" wären mit den Merkmalen korrespondierende Eigenschaftsnamen.

Das Ergebnis analoger Untersuchungen bei frikativen Dauerkonsonanten wird vorgestellt. Es wird gefunden, dass die Frequenzlage einer niederfrequenten Geräuschkante, die Tonheitslage eines empfundenen Geräuschschwerpunktes und die Steilheit einer niederfrequenten Geräuschflanke in signifikanter Weise mit der Zuordnung von Frikativlautklassen zusammenhängen. Hinterlegbare Eigenschaftsnamen wären etwa "Tonhöhe" und "Schärfe".

Literatur

Adam, N., F. Blutner and W. Tscheschner (1974): "A Perception model for processing speech", Proc. Speech Communication Seminar, Stokholm, 339-348



ZUR REALISATION UND PERZEPTION VOKALISCHER /r/-ALLOPHONE DES DEUTSCHEN

Horst Ulbrich, Städtisches Krankenhaus Berlin-Prenzlauer Berg, Phoniatrie Abt., DDR 1055 Berlin

Die r-Realisationsformen des Deutschen sind äusserst heterogen. Sie können nicht nur an verschiedenen Stellen und mit verschiedenen Artikulatoren gebildet werden, sondern sie weichen auch in ihrem Gehöreindruck mehr oder weniger voneinander ab. Sie besitzen auditiv-perzeptorisch nicht nur die Merkmale von vibrierenden und frikativen, sondern auch die von vokalischen Lauten. Sie können sowohl nach dem artikulierenden Organ oder nach der Artikulationsstelle wie auch nach dem Gehöreindruck bezeichnet werden. In der Standardaussprache des Deutschen werden vibrierende und frikative (volle) wie auch vokalische (reduzierte) r-Realisationsformen gesprochen. In einem Corpus von über 10000 - mit Hilfe eines Segmentiergerätes instrumentalphonetisch-auditiv untersuchten - r-Realisationen in natürlich gesprochener Sprache wurden neben 47% voll realisierten r-Varianten (davon 8% vibrierende und 39% frikative r-Realisationsformen) 41% vokalische r-Varianten registriert. Darüber hinaus wurden 9% Elidierungen und 3% unbestimmbare (indifferente) r-Realisationsformen ermittelt. Sowohl die sprechüblich gewordenen Realisationen vokalischer /r/-Allophone nach Langvokalen und in der Phonemsequenz /er/ in Vorsilben und Endungen als auch die Interpretation entsprechender Realisationsformen durch eine Reihe von Abhörern werden kurz erläutert. Darüber hinaus werden an Hand ausgewählter Beispiele einige elektroakustische Registrierungen (Sonagramme) besprochen.

Literatur

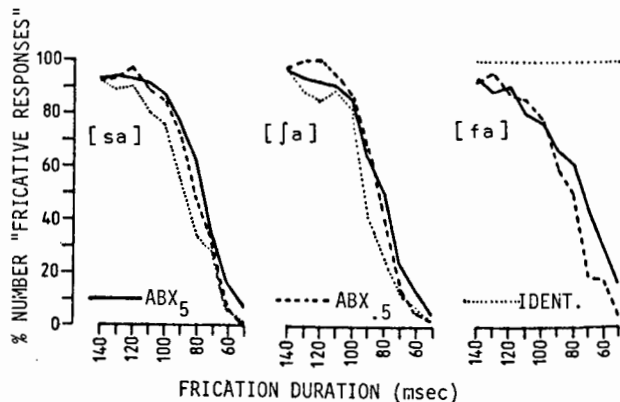
- Krech, H. et al. (Hrsg.) (1. 1964, 4. 1974): Wörterbuch der deutschen Aussprache, Leipzig: VEB Bibliographisches Institut.
- Meyer-Eppler, W. (1954): "Zur Spektralstruktur der /r/-Allophone des Deutschen", Akustika 1, 247-250.
- Ulbrich, H. (1972): Instrumentalphonetisch-auditive R-Untersuchungen im Deutschen, Berlin: Akademie.

## AUDITORY AND PHONETIC PROCESSING OF ITALIAN VOICELESS FRICATIVES SHORTENED IN DURATION

K. Vagges, G.M. Pelamatti and F.E. Ferrero, Centro di Studio per le Ricerche di Fonetica (C.N.R.), Padova, Italy

In order to investigate the possibility of a coexistence of the auditory and phonetic mode of processing for synthetic voiceless Italian fricatives, we performed two ABX discrimination tests with stimuli varying in frication duration, one with 5 sec and another with .5 sec interstimuli interval. The latter should induce the subjects to discriminate the stimuli on the basis of the variable acoustic characteristic (frication duration) rather than on the basis of the phonetic classification shown in an identification test (Ferrero et al. 1978).

The discrimination functions obtained for the three syllables are similar (see figure). Only the discrimination functions of syllables [sa] and [ʃa], compared to the results obtained in the identification test are similar, supporting the conclusion of Ferrero et al. (1978) according to whom, these fricatives are processed phonetically. While the identification task for [fa] seemed to be based on a phonetic analysis, the discrimination task seems to induce an auditory analysis. These results may suggest that the processing of the fricatives involves both auditory and phonetic stages of information processing. The same conclusions are drawn comparing the performance of the subjects for syllable [fa] in the discrimination tests with .5 sec and 5 sec interstimuli interval.



#### Reference

- Ferrero, F., G. Pelamatti, and K. Vagges (1978): "Perceptual category shift of Italian fricatives as a function of duration shortening" submitted for publication in Frontiers of Speech Communication Research, B. Lindblom and S. Ohman (eds.), London: Academic Press.

## THE REFLEX THEORY OF SPEECH PERCEPTION

Jia-lu Zhang, Institute of Physics, Academia Sinica

The role played by semantics and syntax in speech perception and the design of automatic speech recognition systems have attracted much attention. The important role of syllable formation rules is considered and it is pointed out that the syllable formation rules are just what Fletcher calls influence "X" (1953, 286), which appears from our establishment of the statistical relation between syllable and phoneme identification.

Subjects

The perceptual confusion among Chinese consonants was investigated under 18 different transmission conditions, and some comparative investigations were made between Chinese and English (Miller and Nicely, 1955) and Japanese (Nagai et al., 1956). It is shown that: 1. Manner of articulation has priority over place of articulation in speech perception, 2. the social characteristics, i.e. linguistic structure as a social convention, strongly influence speech perception, and the relative importance of each distinctive feature is different in different languages, 3. the syllable structure of Chinese helps in identifying the place of articulation and therefore the correct identification of syllables is increased.

Conclusion

Speech perception is a unitary process that is based on the physical characteristics combined with the social (= structural) characteristics of speech. In this process, all factors in the speech signal are utilized by listeners, the factors playing different roles under different listening conditions and at different stages of speech perception.

References

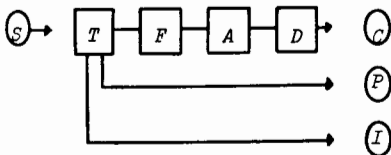
- Fletcher, H. (1953): Speech and hearing in communication, New York: Van Nostland.
- Miller, G.A. and D.E. Nicely (1955): "An analysis of perceptual confusions among some English consonants", JASA 27, 338-352.
- Nagai, K. et al. (1956): "Analysis of phonemes by articulation tests", Journal of the Acoustical Society of Japan 12, 148-154.

PRELIMINARY STUDIES FOR THE AUTOMATIC RECOGNITION OF GERMAN  
SPEECH SOUNDS

Antonio Almeida, Institute of Phonetics, University of Cologne

The intention of the current work is to lay the fundamentals of a system which later on should be able to give a broad transcription of German spoken utterances. For this purpose and before we proceeded to design the system described below, we treated the acoustic data from vowels of ten female subjects by means of discriminant analysis, the results of this research being published in Almeida (in print). We hope to get on to the analysis of German consonants on the same basis before we meet in Copenhagen.

The provisory architecture of the system has the following components:



1) Signal input  $S$ ; 2) Periodicity detection or adaptive time window  $T$ ; 3) Fourier analysis  $F$ ; 4) Data reduction by averaging spectra  $A$ ; 5) Discrimination of sounds  $D$ ; 6) Output of a nonlinguistic segmental chain  $C$ , of a pitch curve  $P$ , and of an intensity curve  $I$ .

At the moment we are making efforts to link the different FORTRAN routines for points 1 to 4 into a coherent system, moreover a real time Fourier analysis is at the verge of completion. We hope to begin with the implementation of the discriminatory component soon.

As stated above, the segmental output will be nonlinguistic, that is to say there will be as many identified segments as average spectra and so a 1 s utterance will have 33 segments if you use a 10 ms window and average on three windows. The conversion of these nonlinguistic phonetic data into broad transcription will be an enterprise for coming years.

Reference

Almeida, A. (in print): Nasalitatsdetektion und Vokalerkennung (=Forum Phonetikum 17), Hamburg: Buske-Verlag

## CALCUL ANALOGIQUE DE LA FREQUENCE DU FONDAMENTAL

Charles Berthomier, Département de Recherches Linguistiques,  
Université Paris VII, 2 Place Jussieu, 75005 Paris, France

On décrit dans cet article une méthode de calcul de la fréquence fondamentale d'un signal de parole. Cette fréquence est calculée de manière analogique en associant au signal de parole préalablement filtré un couple de signaux en quadrature obtenus au moyen d'un réseau déphaseur. Ce couple de signaux peut être considéré comme décrivant la trajectoire d'un point dans un plan, la fréquence calculée étant, à un facteur  $2\pi$  près, la vitesse angulaire de rotation de ce point, et la distance à l'origine étant l'amplitude du signal. On donne trois exemples de résultats obtenus par cette méthode dont l'intérêt réside en particulier dans la rapidité du calcul.

## LES PROPRIETES ACOUSTIQUES DE / j, q, w, l, r / EN FRANCAIS

Michel Chafcouloff, Institut de Phonétique, Aix-en-Provence

Certains sons du langage qui posent de nombreux problèmes aussi bien du point de vue de leur terminologie que de leur description ou de leur statut n'ont pas encore été l'objet d'une étude exhaustive en français. Dans le présent travail, nous présentons les premiers résultats d'une analyse acoustique qui porte sur l'examen des trois paramètres - fréquence, intensité, durée. Des mots comprenant les sons /j,q,w,l,r/ en position intervocalique accentuée et en contexte vocalique /i,y,a,u/ ont été enregistrés et soumis à une analyse spectrographique.

1) En ce qui concerne les caractéristiques spectrales, la structure formantique de /j,q,w/ n'est sujette qu'à des variations minimales contrairement à celle de /l,r/ qui est fortement sensible aux effets de coarticulation occasionnés par le contexte vocalique.

2) Les différences d'énergie globale entre /j,q,w,l/ et les voyelles adjacentes sont dans l'ensemble réduites alors que celles de /r/ sont beaucoup plus nettes. De plus, l'examen des courbes de F<sub>0</sub> révèle des variations microprosodiques assez importantes à propos de /l/.

3) Du point de vue temporel, alors que /l/ se caractérise par la durée de la tenue, /j,q,w,r/ se distinguent par celle des transitions. Il existe une différence significative entre la durée des transitions initiales et des transitions finales, ces dernières étant toujours plus longues que les premières.

Les premiers résultats de cette analyse ainsi que certaines divergences constatées à propos des données présentées par Delattre montrent à l'évidence que les travaux préliminaires de ce dernier doivent être poursuivis et approfondis. La recherche de nouveaux indices et leur évaluation perceptuelle devrait permettre:

- 1) d'améliorer de façon appréciable la qualité auditive de la parole de synthèse.
- 2) d'aboutir à une classification cohérente de ces sons.
- 3) de définir un statut linguistique qui rende compte de la réalité phonétique.

## CROSS-LINGUISTIC NORMALIZATION

Sandra Ferrari Disner, Department of Linguistics, University of California,  
Los Angeles, California 90024 USA

This paper reviews some of the algorithms for vowel normalization that have been proposed in the literature (Gerstman 1968, Harshman 1970, Lobanov 1971, Nearey 1977) and evaluates them on the basis of their ability to reduce the variance between speakers. It also examines the suitability of each for use in cross-linguistic or dialect studies. Assuming that the published observations of phoneticians are valid indications of the relative quality of vowels in different languages, then a good normalization procedure should not introduce spurious trends into the data. The more highly valued of two normalization procedures is the one which removes more of the variance from the data without appreciably altering the vowel patterns in the languages under study.

Data sets from six Germanic languages--Danish, Dutch, English, German, Norwegian, and Swedish--are utilized in this study. All are taken from published sources. Only the frequencies of the first three formants are available in all of the data sets; consequently, the present investigation is limited to those normalization procedures which utilize these parameters only.

It is concluded that no one normalization procedure is consistently better than others at removing the inter-speaker variance. Some languages are best normalized by one procedure, others by another procedure. The Harshman PARAFAC procedure is least efficient in removing the variance, but it is the only one which does not introduce procedural artifacts into the data. Because it does not depend on the formant means or standard deviations--which vary from language to language--as correction factors, the PARAFAC procedure is best suited to cross-linguistic comparisons.

References

- Gerstman, L.H. (1968): "Classification of self-normalized vowels", IEEE Trans. Audio Electroacoust. AU-16, 78-80.
- Harshman, R. (1970): "PARAFAC: Models and conditions for an 'explanatory' multi-modal factor analysis", Working Papers in Phonetics No. 16, Phonetics Lab, UCLA.
- Lobanov, B.M. (1971): "Classification of Russian vowels spoken by different speakers", J. Acoust. Soc. Am. 49, 606-608.
- Nearey, T. (1977): Phonetic feature systems for vowels. Unpublished doctoral dissertation, University of Connecticut, Storrs.

## VOCAL TRACT THEORY AND BOUNDARY EFFECTS

Gunnar Fant, Dept. of Speech Communication, Royal Institute of Technology (KTH), S-100 44 Stockholm 70, Sweden

Acoustic theory of speech production can be approached on various levels of ambition. The lowest one is to work with models for relating essentials of the formant pattern to a vocal tract model specified by a few parameters. This is the most common approach and is largely directed to the study of vowels. However, such models are less capable of handling absolute values than relational patterns. The next level of ambition is to gain a more profound insight in the actual cavity configurations within the vocal tract including details and overall constraints, consonant articulations, nasal cavity, cavity wall effects, radiation. A third level of ambition is to handle the aerodynamics of the voicing mechanism and of unvoiced sounds so as to enable a proper separation of source and filter characteristics, e.g. for the estimation of the glottis impedance and how the subglottal system affects the speech. At this level of analysis we need to consider second order effects in the analysis of rapidly changing impedance structures. Such effects could also have significance in dealing with rapidly opening or closing of the supraglottal tract. Formant bandwidths are to a considerable extent influenced by vocal tract "boundary" conditions. Of special interest is the temporal modulation of formant bandwidths by the glottal opening and closing within a voice period. The dependency of this modulation on voice register and vowel category will be discussed. Vowels with pharyngeal narrowing are especially sensitive to this damping which can be seen in the oscillogram as a truncation of the signal in the glottal open period.

Literature references appear in an expanded version of this summary.



L'INFLUENCE DU COUPLAGE ACOUSTIQUE LARYNX - CONDUIT VOCAL SUR LA  
FREQUENCE FONDAMENTALE DES VOYELLES. UNE SIMULATION

Bernard Guérin, E.N.S.E.R.G., Louis-Jean Boë, Institut de Phonétique de Grenoble, Roland Lancià, E.N.S.E.R.G., 23, avenue des Martyrs, Grenoble

Dans la parole naturelle, des différences significatives entre les moyennes de la fréquence fondamentale  $F_0$  des voyelles ont été relevées depuis longtemps. Elles se situent entre 4 et 25 Hz et varient peu d'une langue à l'autre: ce sont les voyelles fermées qui ont les fréquences les plus élevées. Pour expliquer ce phénomène, deux hypothèses ont été jusqu'ici retenues: l'influence du couplage acoustique source-conduit vocal et l'interaction physiologique entre la position de la masse de la langue et la tension des cordes vocales.

Les premières évaluations des impédances acoustiques du larynx et du tractus ont fait apparaître qu'une interaction non négligeable pouvait se produire. La simulation permet d'évaluer directement ce phénomène. De nombreuses études ont montré la bonne adéquation du modèle à deux masses, proposé en 1968 par Ishizaka & Matsudaira, malgré les simplifications introduites dans son fonctionnement et ses commandes ( $P_g$  la pression subglottique et  $Q$  un paramètre qui rend compte de la tension passive). Par ailleurs, les derniers travaux de Mrayati & al. ont permis de chiffrer l'impédance d'entrée du conduit vocal, compte-tenu de l'estimation des pertes.

L'étude présentée ici concerne la fréquence de vibration des cordes vocales dans le cas des voyelles orales. Dans un premier temps le problème a été abordé sur un plan théorique. Ont été envisagés les cas où la charge que représente le conduit vocal est capacitive, inductive ou résistive: il est ainsi possible de séparer les effets de chacun des éléments de l'impédance d'entrée sur  $F_0$ . Ensuite, le couplage a été simulé pour différentes valeurs de  $P_g$  (6 et 8 cm d' $H_2O$ ) et de  $Q$  (1, 1.5 et 2.5) et pour les voyelles du français [i y e ø ε œ a o u]. Les résultats montrent que pour  $F_0$  voisin de 120 Hz, le couplage introduit des variations de l'ordre de 8 Hz;  $F_0$  est maximale pour les voyelles ouvertes et minimale pour les voyelles fermées. Ces observations, qui vont dans le sens de l'étude théorique, mais qui sont contraires à celles que l'on observe dans la parole naturelle, tendent à montrer que l'hypothèse du couplage acoustique ne peut être retenue.

Référence: Ishizaka, K. et M. Matsudaira (1968): "What makes vocal cords vibrate", Proc. 6th International Congress on Acoustics, B 13.

## VOWEL ANALYSIS WITH LINEAR PREDICTION

T. de Graaf, Institute of Phonetic Sciences  
University of Groningen, Netherlands

The autocorrelation method of linear prediction is used in order to determine the first four formant frequencies of a number of consecutive speech segments that represent a specific vowel phoneme. This method is able to supply a pattern of formants given as functions of time, which characterize the particular vowel or diphthong. It has been applied to the study of vowel systems of regional languages in the Netherlands, in particular Frisian.

We find some characteristic acoustic features for the Frisian diphthongs, which can be divided into 5 closing diphthongs  $\epsilon i$ ,  $a i$ ,  $\delta i$ ,  $\upsilon u$  and  $\Lambda \ddot{u}$  and 6 opening diphthongs  $i \epsilon$ ,  $\ddot{u} \epsilon$ ,  $i \epsilon$ ,  $\Lambda \epsilon$ ,  $u \epsilon$ ,  $\delta \epsilon$ . Many of these diphthongs are characterized by a short transition segment between an initial and a final stationary part. The formant values F1 and F2 for these stationary parts are obtained as acoustic parameters determining these diphthongs. For several diphthongs we find that the value of F1 and F2 for the first or the last stationary part can differ considerably from the value which belongs to the short vowel representing this part of the diphthong in its phonetic notation.

The opening diphthongs show the property of breaking: an interchange into a rising diphthong with other acoustic parameters, that are also measured and compared to the parameters of the original falling diphthong. Due to language interference on the phonetic level the acoustic manifestation of these phonemes can be changed under the influence of the Dutch language.

In order to study these phenomena in detail a further acoustic analysis is made of a large sample of speech sounds pronounced by different persons (Frisian or non Frisian) in various contexts. Results illustrating the acoustic properties of the Frisian phoneme system and acoustic data related to the process of language interference will be presented in August 1979 at the Congress of Phonetic Sciences in Copenhagen.

UN OUTIL EXPERIMENTAL POUR LE DECODAGE ACOUSTICO-PHONETIQUE  
DE LA PAROLE CONTINUE

Jean-Paul Haton et Claude Sanchez, Equipe de Traitement du Signal et Reconnaissance de Formes, Centre de Recherche en Informatique de Nancy, Université de Nancy 1, C.O. 140, 54037 Nancy, France

Cet article présente une approche de type Intelligence Artificielle de la reconnaissance acoustico-phonétique de la parole continue, dans le cadre du système général MYRTILLE II de compréhension de la parole continue, actuellement en cours de conception.

La segmentation de la parole en unités phonémiques est effectuée par calcul d'une fonction de différents paramètres acoustiques (intensité, taux de passages par zéro, longueur curviligne du signal, etc...). Ceci permet d'affecter un score aux frontières obtenues, en vue d'une remise en cause ultérieure en cours de traitement.

Pour construire le décodeur phonétique capable de reconnaître ces segments, l'utilisateur définit une batterie de processeurs qui prennent en compte un ou plusieurs traits ou indices phonétiques (voisement, formants, énergies dans certaines bandes de fréquence, etc...). En sortie, ces processeurs fournissent des indications portant soit sur le type de phonème étudié soit sur le processeur à mettre en oeuvre pour poursuivre l'identification. Le système définit la configuration optimale d'association de ces processeurs, sous forme d'une structure arborescente. Ce système est ainsi capable d'engendrer des systèmes de reconnaissance phonémique; il s'avère être très utile pour tester la validité en reconnaissance de différents traits et indices ainsi que de différentes stratégies de reconnaissance.

Les résultats obtenus à partir de différentes implantations de systèmes sont présentés et discutés.

PITCH DETERMINATION OF SPEECH SIGNALS BY NONLINEAR DIGITAL  
FILTERING

Wolfgang Hess, L.f. Datenverarbeitung, TU München, W. Germany

Pitch determination can be done in many ways. In the time domain, the first harmonic can be enhanced by low-pass filtering, or the temporal structure of the signal can be changed in such a way that periodicity is easily detected. Pitch detectors of this type, however, get into trouble when the first harmonic is attenuated or missing. To overcome this problem, the first harmonic must be reconstructed by nonlinear distortion. To study this effect, several nonlinear functions (NLFs) were examined in order to select one that can be applied to any signal within the whole range of pitch. No function, however, meets this requirement with optimal performance. Thus a combination of three NLFs (odd, even, and SSB) was selected, giving a good approximation to the ideal case. Using these NLFs, a given pitch detector (Hess, 1976) has been modified so as to make it independent of the type of input signal. The signal is first simultaneously processed by the three NLFs. The subsequent linear filtering steps represent a crude approximation to the inverse filters. A low-pass filter removes the higher formants (separately for each of the 3 channels). Then  $F_1$  or, for the even NLF, the dominating frequency resulting from the filtering after distortion is determined in each channel. The subsequent adaptive band-stop filter removes this dominating frequency; its zero, however, is commonly adjusted for the 3 channels to the highest of the "formant" frequencies actually measured. This ensures that the first harmonic is never suppressed, even when it coincides with  $F_1$ . Hence, the output signal of the band-stop filter contains a strong first harmonic at least in one channel. Deriving preliminary pitch period boundaries (markers) in each channel, and checking the regularity of these markers during short intervals (25 ms), the algorithm selects the appropriate channel for final processing. In a preliminary test, the algorithm showed good performance for undistorted as well as for band-limited signals within a range of fundamental frequencies from 70 to 500 Hz.

Reference

Hess, W. (1976): "An algorithm for time-domain pitch period determination", IEEE Intern. Conf. Acoust., Speech, and Signal Processing, Philadelphia PA (ICASSP-76), 322-325.

## METHOD FOR IDENTIFYING TALKERS FROM ACOUSTIC SPEECH ANALYSIS

Harry Hollien, Charles C. Johnson, and Wojciech Majewski  
IASCP, University of Florida, Gainesville, FL USA and  
Wroclaw Technical University, Wroclaw, Poland

A four vector semi-automatic speaker identification system (SAUSI) has been described (Proc. IEEE: ASSP, 1977, 768-771); the SYSTEM NOW EMPLOYS SIX MAJOR VECTORS. In order to evaluate the validity of a system such as this one, the vectors must be tested by a large number of protocols both singly and in groups. In order to permit such testing, we have generated a very large database grouped into three general categories: laboratory, field simulation and field; they include: 1) normal speech produced in two languages by large populations of subjects, 2) laboratory quality speech produced as a function of stress and disguise, 3) high quality "field" speech (transmitted by radio) produced under stress, 4) speech produced by talkers of different dialects plus dialect imitators, 5) speech produced over telephone links -- included is normal speech and a variety of controlled disguises and 6) simu-crimes recorded in the field. Virtually all of the over 1000 sample-sets are of male talkers; however the category No. 5 includes 25 women.

The six vectors currently utilized are generated from 11-60 parameters each; they include: 1) fundamental frequency (17-25 parameters), 2) power spectra (11-23 parameters), 3) vowel formants (32-45 parameters), 4) phoneme structure (60 parameters), 5) vocal jitter (variable parameters), 6) temporal features (15-24 parameters). The first two and the last vectors have been subjected to considerable laboratory analysis -- for both large and small populations and under both ideal and distorted speech conditions. Some testing of the other vectors and of combinations of vectors also has been carried out. The results have been encouraging and experiments currently are under way evaluating the identification power of the combined vectors.

ASPECTS ACOUSTIQUES DE LA VOIX DE TRANSEXUELS: TON FONDAMENTAL  
ET FREQUENCES FORMANTIELLES

Benoit Jacques et Clau Désautels, Département de Linguistique,  
Université du Québec à Montréal, Montréal, Québec, Canada

Le but de cette recherche était de déterminer dans quelle mesure des personnes qui changeaient de sexe pouvaient acquérir une voix dont le registre correspondait à leur nouvelle identité sociale.

Déroulement de la recherche

Notre corpus était constitué d'enregistrements de dix (10) transsexuels, 5 femelles génétiques qui désiraient appartenir au sexe masculin et 5 mâles génétiques qui désiraient passer au sexe féminin. Une analyse sonographique de ces enregistrements a été faite; nous avons mesuré le Fo et les formants des voyelles produites par tous les sujets.

Résultats

Certains candidats n'arrivent pas à acquérir le registre de voix correspondant à leur nouveau sexe et il y a peu de corrélation entre le traitement hormonal et la fréquence du Fo. Cependant on a constaté que certains candidats suppléent à cette difficulté par un comportement articulatoire qui a pour effet d'altérer le timbre des voyelles. Ce comportement apparaît plus déterminant sur la perception auditive d'une voix comme étant féminine ou masculine que la fréquence du Fo.

Bibliographie

- Brown jr, W.S. et S.H. Feinstein (1977): "Speaker sex identification utilising a constant laryngeal source", Folia Ph. 29,3.
- Coleman, Ralph O. (1972): "The perception of maleness and femaleness in the voice and its relationship to vowel formant frequencies", in Proc.Phon. 7, The Hague: Mouton.
- Coleman, Ralph O. (1976): "A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness in the voice", JSHR 19, 1.
- Tarnaud, J. (1961): Traité pratique de phonologie et de phoniatrie, Paris: Librairie Maloine.

## STATISTICAL CLASSIFICATION OF POLISH FRICATIVE CONSONANTS BASED ON THEIR SPECTRAL FEATURES

Wiktor Jassem, Acoustic-Phonetics Research Unit, Polish Academy of Sciences, Poznań

Spectra of a total of 1035 fricatives spoken in nonsense-words by 3 voices were specified in three different ways and the following quantitative features were used for their classification: coefficients of the terms in polynomials describing the spectral envelope, partial areas under the envelope and centres of gravity. According to the kind of feature, between 1 and 12 variables were used in a statistical model which divided the feature space into classification regions - either one for each phoneme or one for each variant. With a maximum of 12 variables a 100% correct classification could be obtained by applying quadratic statistical discriminant functions. Under specific conditions, with no more than 4 variables, 99% correct classification could be achieved.

The analysis was performed by connecting the Sona-Graph via an A/D converter to a minicomputer and the mathematical operations, including the decision-taking were carried out in a larger general-purpose computer. Under less-than-optimum conditions, /s,z,x/ gave slightly better results than the other phonemes.

The methods using centres of gravity and those using partial areas under the spectral envelope appear more satisfactory than those employing polynomials.

As speech sounds can be correctly classified by using instrumental (acoustical) analysis and mathematical data processing, it is suggested that phonemes and their variants may be regarded as objective, physically distinct entities.

## PHONETIC EXPLANATIONS FOR DEVOICING OF HIGH VOWELS

Hector R. Javkin, Phonetics Laboratory, University of California, Los Angeles

It has been well established (Greenberg 1966, Jaeger 1978) that high vowels devoice more frequently than low vowels. Ohala (1975) suggested two explanations on the basis of a model of speech aerodynamics. The model predicted that oral air pressures would be higher for high than for low vowels, thus reducing the pressure drop across the glottis necessary for voicing; and that air velocity would be greater at the place of maximum constriction for high than for low vowels, resulting in more noticeable frication. A further hypothesis is that the transfer function of the vocal tract results in greater fricative noise for high than for low vowels. Measurements from one speaker suggest that the pressure differences between high and low vowels cannot be the explanation, since those pressures, averaged over three environments, are essentially equal:

/i/	.40 cm H <sub>2</sub> O
/u/	.54 cm H <sub>2</sub> O
/ɔ/	.51 cm H <sub>2</sub> O
/æ/	.43 cm H <sub>2</sub> O

The hypothesis that the transfer function is responsible for the greater noisiness of high vowels was tested with a computer vocal tract model which produced random noise at the place of maximum constriction for three modeled Russian vowels. The output of the model was subjected to a Fourier analysis, which did not yield relevant differences in fricative amplitude. It is clear from this study that the explanation must lie in the narrower constriction and greater air velocity for high than for low vowels.

#### References

- Greenberg, J. (1966): "Synchronic and diachronic universals in phonology", Lg. 42:508-17.
- Jaeger, J. (1978): "Speech aerodynamics and phonological universals", Proc. Berkeley Ling. Soc. 4:311-29.
- Ohala, J. (1975): "A mathematical model of speech aerodynamics", in Speech Communication. Proceedings of the Speech Communication Seminar, Stockholm, Aug. 1-3, 1974. Vol. 2, pp. 65-72.



SUR QUELQUES INDICES ACOUSTIQUES DES SONS STABLES DU FRANCAIS  
EMIS PAR PLUSIEURS LOCUTEURS

J.S. Liénard, Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI) - B.P. 30, 91406 Orsay, France

Pour aborder le problème de la variabilité des sons de la parole selon le locuteur et le type de voix nous avons étudié les spectres de 18 sons stables (voyelles orales et nasales, consonnes constrictives sourdes, occlusion voisée, bruit d'ambiance) provenant de l'enregistrement de 2 enfants, 2 femmes et 4 hommes dont l'un utilise également la voix de fausset et la voix chuchotée.

La détection et l'identification des formants conduisant à de nombreuses erreurs, nous avons recherché des indices acoustiques plus simples. Une première expérimentation a permis de définir deux indices rendant compte des dimensions grave-aigu et compact-diffus.

Dans une seconde expérimentation nous avons considéré des paramètres provenant d'une analyse fréquentielle très grossière. Le plus significatif d'entre eux correspond à la courbure du spectre très sévèrement lissé, aux environs de 900 Hz. Il permet à lui seul d'opposer deux à deux les phonèmes du corpus avec 49% de chances de succès si l'on tolère un maximum de 5% d'erreurs de classification, et 73% de succès si l'on tolère un maximum de 32% d'erreur.

En complétant cette classification par l'utilisation de trois autres paramètres calculés de la même manière mais choisis à d'autres fréquences, les taux de succès passent à 57% (avec un maximum de 5% d'erreur) et 93% (avec un maximum de 32% d'erreur). Les oppositions non résolues correspondent à des sons très voisins comme [a]-[ã], [o]-[ɔ], dont on ne peut même affirmer qu'ils aient été clairement distingués par tous les locuteurs du corpus.

Ces expérimentations sont encore sommaires. Elles permettent cependant de remettre en question la pertinence de la notion de formant, au moins en ce qui concerne la perception des sons stables du français. Un autre aspect intéressant est la progressivité des valeurs prises par les indices acoustiques, progressivité dont les théories binaires des traits distinctifs font un mauvais usage.

## INTONATION: ANALYSE ACOUSTIQUE ET PERCEPTIVE DU PORTUGAIS

M. Raquel Delgado Martins, Laboratório de Fonética, Faculdade de Letras da Universidade de Lisboa - Portugal

Cet article présente les premiers résultats d'une recherche portant sur les rapports de certains indices acoustiques et la perception de l'intonation. Ce travail prétend apporter une contribution à la notion de "réalité de la représentation phonétique" (Chomsky, 1968, 24) et à d'autres notions aussi controversées que celles même d'intonation, d'accent ou de syllabe (Ladefoged, 1975), à partir des faits d'accent en Portugais.

MATERIEL

L'enregistrement de dix phrases du Portugais a été utilisé pour ce travail. Les valeurs des indices de fréquence fondamentale, d'intensité, d'énergie et de durée ont été recueillies et ordonnées en fonction des valeurs pour l'ensemble de la phrase. Cet enregistrement a été également présenté en test à 32 sujets portugais, à qui il était demandé, au cours de plusieurs auditions d'une même phrase, d'attribuer un degré hiérarchisé d'accent à chaque syllabe en fonction de la totalité de la phrase. Les données de ce test sont comparées aux résultats du traitement des indices acoustiques.

CONCLUSION

Cette procédure d'analyse permet de tirer certaines conclusions sur l'organisation de certains indices acoustiques et leur évaluation perceptive qui viennent confirmer les hypothèses posées à ce sujet au niveau de l'accent de mot dans un travail antérieur (Delgado Martins, 1977). Les résultats du test sont significatifs quant à une effective perception de l'intonation et montrent l'importance du rapport énergie/durée et de la qualité phonologique pour la perception.

Références

- Chomsky, N. and M. Halle (1968): The Sound Pattern of English, New York: Harper and Row.
- Delgado Martins, M.R. (1977): Aspects de l'Accent en Portugais, Thèse de Doctorat non-publiée, Strasbourg.
- Ladefoged, P. (1975): A Course in Phonetics, New York: Harcourt, Brace and Janovich.

SEGMENTATION ET RECONNAISSANCE ACOUSTIQUE PHONETIQUE  
DE LA PAROLE CONTINUE

G. Mercier, C.N.E.T., route de Trégastel, 22301 Lannion, France

Dans cette communication, on présente les principaux paramètres acoustiques utilisés par l'analyseur phonétique du système K.E.A.L. de reconnaissance de la parole continue.

Description des paramètres

Les paramètres de base de cet analyseur phonétique sont les sorties d'un vocodeur à canaux (14 filtres) et du détecteur de pitch. A partir de ce spectre on calcule d'autres paramètres tels que l'énergie  $E(t)$  toutes les 13,3 ms, la dérivée  $P(t)$  du signal, le centre de gravité fréquentiel  $G(t)$ , la variance du spectre autour de sa valeur moyenne, la position des maxima du spectre et leurs variations au cours du temps.

Procédures de segmentation et d'identification

A partir de ces paramètres, l'analyseur phonétique détecte le début et la fin de parole et segmente la parole en syllabes.

On utilise ensuite une procédure hiérarchisée et un ensemble de règles contextuelles qui permettent de séparer les voyelles des consonnes, de détecter selon les cas des segments voisés ou non voisés, des segments fricatifs, plosifs, nasals ou liquides, ou de ne pas prendre de décision lorsque les marques acoustiques ne sont pas suffisantes.

A l'issue de cette procédure, le programme essaie d'identifier à l'intérieur de chaque classe le phonème prononcé à l'aide de fonctions de séparation linéaires dont les coefficients sont préalablement calculés pendant une phase d'apprentissage.

Résultats et conclusions

La communication elle-même présente les résultats obtenus à chaque niveau d'analyse, essaie d'expliquer les causes d'erreurs et suggère quelques solutions permettant d'y remédier.

Références

Gresser, J.Y. et G. Mercier (1975): "Automatic segmentation of speech into syllabic and phonemic units. Application to French words and utterances", in G. Fant and M.A.A. Tatham (éds.): Auditory Analysis and Perception of Speech, 359-382, London: Academic Press.

Mercier, G. (1978): "Evaluation des indices acoustiques utilisés dans l'analyseur phonétique du système K.E.A.L.", 9<sup>es</sup> Journées d'étude sur la parole, Lannion, 31 mai - 2 juin, 321-342.

ETAGE PHONÉTIQUE D'UN SYSTÈME DE RECONNAISSANCE ET DE  
COMPREHENSION DE LA PAROLE CONTINUE

G. Perennou et J. Caelen, Laboratoire C.E.R.F.I.A., Université  
P. Sabatier, Toulouse, France

Dans un système de reconnaissance de la parole continue, toutes les informations acoustiques doivent être utilisées si l'on ne veut pas alourdir les étapes linguistiques. Par ailleurs il apparaît difficile d'opérer en catégorisations successives: phonèmes, syllabes, mots, ou à l'inverse, hypothèses, phrases, mots, syllabes, phonèmes. En effet les indices acoustiques sont parfois trop fragiles pour autoriser la construction d'unités linguistiques et, d'autre part, la méthode descendante qui procède par vérification (et de ce fait demande des indices moins précis) est inadéquate à partir d'un facteur de branchement au-delà de quelques dizaines.

On décrit ici un processus de reconnaissance au niveau phonétique et on y distingue trois étages qui assurent la continuité entre la réalité acoustique du signal et la chaîne phonétique abstraite. L'étage acoustique élabore des données par blocs de 8 ms qui sont les paramètres étudiées  $P_1$ . L'étage suivant est une transition vers le niveau purement phonétique. On y élabore des segments encore acoustiques mais à vocation phonétique dotés des propriétés  $P_2$ .

Enfin, le dernier étage, purement phonétique a pour but de proposer des candidats phonèmes à partir de segments acoustiques qui pourront à ce niveau être amalgamés (cas des explosives sourdes par exemple) redécoupés (cas de segments vocaliques longs et non homogènes). Ces segments acoustiques pourront dans beaucoup de cas correspondre directement à un phonème. Enfin, ils seront laissés en l'état, lorsque des propriétés claires ne permettent pas d'y localiser des phonèmes.

## KAZAKH VOCAL SPEECH AND SPECTRAL CHARACTERISTICS OF VOWELS

S.S. Tatubaev, Alma-Ata, USSR

The work aims at investigating the spectra of vowels in speech and singing. For this purpose we developed a method of analysis, namely the syllabic-matrix system with statistical distributions of vowels and consonants.

The formant characteristics of the vowels in speech and singing were determined. The vowels were sung in different singing registers. In singing there is a change in the vocal tract, whereas in speech this change is less pronounced.

The analysis of the vowels in singing shows that the vowel consists of two timbres: the general timbre, which characterizes this given vowel and whose frequency is below 2500 Hz, and the second timbre which is above 2500 Hz and which is called the "singing formant" part. The singing formant part is connected with such important qualities of the voice as brightness and flight which has different width and different amplitude characteristics.

The presence of the singing formant in our investigated spectra of kazakh-singers shows that the singing formant is characteristic of singing in Kazakh as well and probably depends mainly upon the technology of voice formation. The singing formant does not depend on the type of the voice.

## UN SYSTEME DE DETECTION AUTOMATIQUE DE LA SONIE DES SONS DU LANGAGE

B. Teston, Institut de Phonétique d'Aix-en-Provence

Le système que nous décrivons est un détecteur d'intensité des signaux acoustiques du langage auquel sont appliquées différentes pondérations de manière à obtenir une mesure la plus proche possible de la sensation auditive effectivement perçue.

Ces pondérations tiennent compte:

- des courbes isosoniques
- de la répartition de l'énergie du signal dans le spectre
- de l'effet de masque.

L'appareil est essentiellement constitué par un analyseur de fréquence en temps réel dont les filtres d'analyse ont une progression simulant les bandes critiques de Zwicker (1957).

Un détecteur de voisement permet de faire la distinction entre les signaux voisés et non voisés. Si le signal est constitué par des sons voisés, le calcul de la sonie est alors effectué au moyen de la méthode proposée par Rossi (1971). La valeur de l'intensité pondérée du signal de parole est obtenue directement en phones toutes les 10 millisecondes. L'exploitation des résultats peut être réalisée au moyen d'un oscillographe enregistreur sur lequel est visualisée la courbe d'intensité comme avec un intensimètre classique auquel l'appareil se substitue. Il évite ainsi un fastidieux et long travail de pondération manuelle à partir d'une courbe d'intensité objective. Le système est également connecté à un ordinateur pour réaliser des traitements particuliers sur les paramètres prosodiques.

L'analyseur du détecteur de sonie peut effectuer des analyses spectrales par octave et par 1/3 d'octave. Il est également possible de pondérer l'intensité et la constante de temps de chaque bande d'analyse.

Il est prévu de faire évoluer l'appareil pour intégrer l'influence de la durée sur la perception des segments acoustiques. Des recherches dans ce sens sont envisagées à court terme dans notre laboratoire, ainsi que des études sur la perception des consonnes non voisées, dont nous comptons également intégrer les résultats pour améliorer le calcul de la sonie de ces éléments.

Références:

- Rossi, M. (1971): "L'intensité spécifique des voyelles", Phonetica 24, 129-161.
- Zwicker, E., G. Flottorp et S.S. Stevens (1957): "Critical bandwidth in loudness summation", JASA 29, 548-557.

## COMMENTS ON THE MYOELASTIC-AERODYNAMIC THEORY OF PHONATION

Ingo R. Titze, Sensory Communication Research Laboratory, Gallaudet College, Washington, D. C. 20002, USA

The myoelastic-aerodynamic theory of phonation has been quantified with computer models of varying complexity during the past decade. Mathematical statements of physical laws were used to simulate air and tissue movement in the larynx, as well as wave propagation in the vocal tract. The feedback mechanism by which oscillation of the vocal folds is produced appears to be a result of pressure distributions which are asymmetric with respect to the medial surface velocity of the tissue. Upward and lateral movement during opening is associated with a substantially different pressure profile than downward and medial movement. This mechanism, when sustained, allows energy to be transferred from the air stream to the tissue. In the simplest one-mass model, the asymmetry usually begins with voice onset transients, but may be sustained in the steady state by the inertial inductance of the air in the glottis, or by the vocal tract input impedance. Since these may vary with the direction of flow acceleration (which in turn varies with medial surface tissue velocity), an asymmetric pressure profile can be maintained. With additional degrees of freedom in tissue movement (multiple-mass or continuum models), the pressure distribution becomes asymmetric primarily as a result of combinations of normal mode displacements.

The fundamental frequency seems to be myoelastically controlled. Theoretical models do not support the view that  $F_0$  is controlled by an effective aerodynamic stiffness. Physiologically and phenomenologically, subglottal pressure does affect  $F_0$ , as has been repeatedly demonstrated experimentally, but it appears to be an amplitude related phenomenon which is governed by nonlinear properties of tissue elasticity. Recent measurements on various tissue layers of the vocal folds, as well as the entire literature on the viscoelasticity of human tissue, confirm that the common exponential stress-strain curves can easily account for the observed frequency-amplitude dependence. The negative Bernoulli pressure, which pulls the vocal folds medially prior to glottal closure, is short-range. Over the entire glottal cycle it resembles a mechanical impulse, which imparts momentum, but has little effect on the natural frequency of oscillation.

## REMARKS ON THE GLOTTALIZATION IN JAPANESE

Sataro Uchita, Department of Foreign Languages, Aichi University of Education, Kariya City, Japan

This paper reports the findings of an acoustic study of the glottalization of the Tonyu dialect in Central Japan, where the isogloss between the Eastern and the Western dialects runs from the north to the south. This dialect still preserves some archaisms (both phonological and morphological). The glottal stop, though used by some speakers, has no phonemic status in most dialects (exception: the Ryukyu dialects). It occurs, however, as a regular phoneme in the utterances for the phonological forms /-qb-, -qd-, -qzj-/ (in phonetic transcription [-ʔb-, -ʔd-, -ʔdʒ-]) in the Tonyu dialect.

#### Method and Materials

The glottalized sounds were studied by means of an electroglottograph and a fundamental frequency meter. Glottis vibrations, pitch contours and duplex-oscillograms were recorded in the form of oscillographic photos. Frequency spectra of consonants and vowels, amplitude, and duration were recorded by a sound spectrograph, and the speech signal was simultaneously recorded on magnetic tape. The speech materials used for the investigation were taken from the basic vocabulary of 200 items, and two female native speakers of the Tonyu dialect supplied the materials.

#### Results

It appears that the glottalization is characterized by a long closure and an abrupt rise of the fundamental frequency. In general, the results accord well with predictions made from the auditory impression.

Table 1. Closure duration in msec

Dialect	Tonyu	Tokyo	Owari
Informant	MT, HY	SU	SM
Word			
/he <sup>1</sup> bi/	-	70	100
/he <sup>1</sup> mbi/	-	-	170
/he <sup>1</sup> qbi/	200	-	-

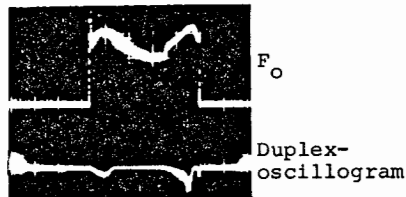


Fig. 1. The pitch contour of /he<sup>1</sup>qbi/ spoken by TM.

#### References

- Hattori, S. (1973): "Japanese Dialects", in *Current Trends in Linguistics* 11, 368-400, Thomas A. Sebeok (ed.), The Hague, Paris: Mouton.
- Nomura, M. (1977): "Remarks on the description and the process of formation of the Tonyu dialect", in *the Bulletin of the Faculty of Letters*, LXXIII, 3-17. Nagoya University, Nagoya.



THE RELEVANT FEATURE, THE NON-CORRELATED PHONEME AND  
THE 'CASE VIDE'

Tsutomu Akamatsu, University of Leeds, England

The notion of the relevant feature is one of those with which functional phonology operates. I propose to discuss just one of the functionalist characteristics of this notion and some consequences that seem to follow from it.

Several scholars have discussed the question of the phonological status of voicelessness for the non-nasal phoneme in the dorsal order in the Dutch consonant system. According to a widespread analysis, the phoneme in question is identified as /k/ and also as a non-correlated phoneme because it lacks a potential partner phoneme which would be /g/ and, moreover, voicelessness which is generally admitted to be non-distinctive for the /k/ (though distinctive for /p/ and /t/ in Dutch) is nevertheless presented as if it were relevant or functional.

However, strict adherence to the notion of the relevant feature with which functional phonology traditionally operates would render the afore-mentioned analysis untenable in that, apart from the fact that this analysis fundamentally violates the defining characteristic of the relevant feature, it would be unjustified to identify the dorsal phoneme concerned as /k/ in the first place and further to envisage only one 'case vide' which would correspond to /g/, i.e. the potential partner of /k/.

## PHONOLOGICAL INTERPRETATION OF NEO-ETYMOLOGIZED PHONEMES

Z.M. Al'mukhamedova, Faculty of History and Philology, Kazan State University, Kazan, USSR

Under neo-etymologization we understand the substitution of one member of a neutralizable opposition by another in the position of maximum differentiation. According to the terminology of the Moscow phonologists, the distinctive unit found in the position of neutralization and absent in the position of maximum differentiation is called a hyperphoneme, and it is transcribed phonologically by two or more symbols, e.g., since there is neutralization between /o/ and /a/ in unstressed syllables, the word for 'dog' sobaka, phonetically [sa<sup>1</sup>baka], is phonologically /s<sup>1</sup>o-a<sup>1</sup>bāka/. The etymologically correct underlying phoneme often appears in the orthography (as in sobaka) and very often also in the pronunciation of alternating forms where the unit is found in the position of maximum differentiation, like acc. vódu /vódu/, pronounced [<sup>1</sup>vodu] 'water', nom. vodá, /v<sup>1</sup>o-a<sup>1</sup>dá/, pronounced [va<sup>1</sup>da]. By neo-etymologization the hyperphoneme is interpreted as representing the other possible phoneme (or one of the other possible phonemes), and this is inserted in the position of maximum differentiation. There is thus a diachronical change of underlying phoneme. This may appear in the literary language, like Tóma, nickname of Tamára, (sometimes both appear like Lóra and Lára, from Larissa). Very often such neo-etymologizations appear in jargon (like lor for laringólog), in children's speech (like [<sup>1</sup>f<sup>1</sup>aki], plur. of flag /fla<sup>1</sup>g-k<sup>1</sup>/, pronounced [f<sup>1</sup>ak] with final devoicing), also in dialectal and in colloquial speech, i.e. in cases where correction from orthography does not so easily take place. The loosening of the semantic relations between cognate forms favours these developments like [<sup>1</sup>votot<sup>1</sup>ka] in dialectal speech (in the orthography vodočka) from [votka] /vó<sup>1</sup>d-t<sup>1</sup>ka/, related etymologically to vodá 'water' (the d has become voiceless before a voiceless consonant, and is interpreted as t).

The existing explanations by factors of morphology or substratum are, in our opinion, not quite correct.

## SIMULATION OF PHONOLOGIES

Thomas D. Arkwright, Defense Language Institute Foreign Language Center, Presidio of Monterey, California 93940, U.S.A.

The object of this presentation is to indicate the existence of a computer system that uses phonological rules as a basis for applying rules to data. This system, PHONOL, can be used by researchers who wish to develop and test a system of phonological rules for a given language.

The main advantage of this tool is that, while it applies rules, the phonologist is not required to learn how to write computer programs. Due to the precision, simplicity and rapidity of PHONOL, the phonologist is enabled to attend to the application of his theoretical knowledge to his chosen problem.

To use PHONOL, the researcher submits a set of rules and a set of base forms to the computer. For each base form PHONOL produces a derived phonetic form, and lists the names of any rules which have applied.

Given results, the researcher would typically change rules or data as needed, and repeat the process in an attempt to produce a consistent set of rules that can account for a representative corpus.

Some secondary aspects of PHONOL are to be noted in this presentation, three of which are:

- (1) the production of all phonetic variants obtainable from a base form, given that one or more rules is optional;
- (2) the automatic checking of phonetic forms against the results expected by the phonologist;
- (3) the use of diverse notational systems (IPA, etc.) based on the printed results produced by PHONOL.

## A MODEL THAT YIELDS ALL ALTERNATIVE PRONUNCIATIONS

Thomas D. Arkwright and Andrew Kerek, Defense Language Institute Foreign Language Center, Presidio of Monterey, California, U.S.A. and Miami University, Oxford, Ohio, U.S.A.

We usually think of phonologies as a means to convert any single base form into a single pronounced form (which is usually the normative form). However, many base forms can be pronounced in more than one way, so it seems that phonologies should be able to produce all observable alternative pronunciations.

Having noted that a fundamental requirement is not met by current phonological theory, this paper presents a phonological model that can produce all observable pronunciation alternatives.

In principle, a phonology that contains  $n$  rules can produce a maximum of  $2^k$  alternative pronunciations from any base form, where  $k$  is the number of rules that are optional. In practice, the number of possible pronounced forms that can be produced is  $2^j$ , where  $j$  is the number of optional rules that can apply to a given base form (as can be determined by an iterative procedure). The observable pronunciations can be found among the  $2^j$  possible pronunciations. Our model produces the set of possible pronunciations, and uses empirically-defined interrule relations to select all observable pronunciation alternatives.

A powerful generalization lies in our finding that the optional/obligatory property of rules is naturally defined by the values along the diagonal of the square matrix that expresses all interrule relations.

This model has been simulated by a computer, so we shall present sample derivations and experimental findings.

## LE SON ET LA TRANSCRIPTION DITE PHONETIQUE

Eric Buyssens, Bruxelles

Malgré la révolution introduite en phonétique par la phonologie, beaucoup d'auteurs continuent à attacher au mot "son" le sens qu'il avait au début du siècle.

Lorsqu'on analyse un graphique comme le spectrographe nous en donne, on constate qu'à l'endroit qui correspond, par exemple, à la voyelle du mot "chat", dans une phrase comme "Où est le chat?", certaines caractéristiques de ce graphique correspondent uniquement au phonème /a/, d'autres font partie de la variante mélodique que l'on appelle l'intonation, d'autres encore résultent de l'intensité avec laquelle le mot est prononcé, d'autres enfin résultent de la rapidité du débit. Réaliser un phonème, c'est-à-dire le prononcer, c'est le combiner à ces trois sortes de caractéristiques qui lui sont étrangères; en outre, c'est le combiner à d'autres caractéristiques qui résultent du sexe du locuteur, de son âge, de son état physiologique ou affectif, etc...

Que l'on place la lettre a entre crochets ou entre barres obliques, elle est incapable de représenter l'une quelconque des caractéristiques qui ne font pas partie du phonème. On peut représenter la hauteur mélodique, l'intensité, le tempo, mais cela se fait autrement que par lettres. Les lettres entre crochets représentent la même chose que les lettres entre barres obliques, à savoir des phonèmes ou leurs variantes, qu'elles soient combinatoires ou libres.

## TOWARDS A GESTALT PHONOLOGY

William M. Christie, Jr., Department of English, University of Arizona, Tucson, Arizona 85721, USA

Raimo Anttila (1977) has offered in outline form a suggestion for what he refers to as a Gestalt Linguistics, an alternative to the atomistic reductionism that has dominated modern linguistics, particularly in the United States, for well over half a century. In this paper I particularize Anttila's proposals to phonology and sketch the outlines of a Gestaltist approach to phonology. The crucial notion that allows the creation of such a model is the complementarity of various approaches in a dynamic tension without requiring of them full integrational compatibility. The approaches to be used are the particle, wave, and field models described by Pike (1959), with the field model particularized as Pike has it to functional field. The method of holding these sometimes conflicting approaches together determines a particular analysis of the goals of phonological theory that limits studies to attempts to gain whatever insights may be available without assuming that the final result will be a comprehensive picture of God's truth about phonology.

References

Anttila, Raimo (1977): "Dynamic fields and linguistic structure:

A proposal for a Gestalt linguistics", Die Sprache 23, 1-10.

Christie, William M. (forthcoming): "Prosodic analysis as Gestalt phonology", Indian Journal of Linguistics.

Pike, Kenneth L. (1959): "Language as particle, wave, and field", The Texas Quarterly 1, 37-54.

## THE ROLE OF TIME IN PHONOLOGICAL REPRESENTATIONS

Richard Coates, School of Social Sciences, University of Sussex  
Brighton, United Kingdom

In most current phonological theories, time is taken simply as a dimension in which phonological representations are performed. Feature systems are frequently classifications based on articulatory space features and acoustic quality features. Time may be mentioned explicitly, e.g. in such features as [ $\pm$  length] , or implicitly, e.g. in such features as [ $\pm$  delayed release]. A demonstration is offered that despite these apparent concessions to time, modern phonologies deal in segment-sized units which have no systematic or principled temporal characterisation.

A view of phonological representations is put forward where time is taken as a constitutive feature, not merely as a dimension of performance, and an attempt is made to interrelate the notions of representation and allegro speech with psychomotor categories. For instance, instructions such as fade the signal, prolong the signal and switch off the signal are built into phonological representations. The hierarchic relations among these instructions and their behaviour in allegro speech are discussed, and used as the basis for the explanation of a significant number of sorts of phonological change. Further, in the light of the historical discussion, severe empirical limitations are imposed upon the nature of phonological representations and upon the type of rules which can be included in phonological systems. For example, the role of the segment is weakened, and intra-word anticipatory rules (phonological) are barred.

## THE DIVERSE ROLES OF GLOTTAL IN PAPUA NEW GUINEA LANGUAGES

Anne M. Cochran, Summer Institute of Linguistics, Ukarumpa via Lae, Papua New Guinea

One definition of glottal stop given by Pei (1966,107) is "a voiceless, fortis, impulsive, unaspirated, simple pressure stop". As a result of definitions such as this, glottal has generally been analysed as a stop consonant in Papua New Guinea despite its diverse roles in various languages.

In this paper, part of Pei's definition is shown to be inadequate for the Gimi language in which there are contrastive glottal stop consonants - fortis versus lenis. While glottal fills the role of a stop consonant in Gimi and some languages of Papua New Guinea, it fills other roles in other languages. It may function as part of a complex consonant phoneme in a few languages while for many it functions as an integral part of the syllable nucleus - i.e. vowel with glottal release, or syllabic nasal with glottal release, not as a stop consonant as has been suggested by many linguists. In other languages glottal occurs only between geminate vowels as a complex syllable nucleus of V?V.

Thus it can be seen that it is unwise for linguists to begin analysis by considering glottal automatically as a stop consonant. Where it should be analysed as a part of a complex nucleus, it has been analysed as occurring as the only stop segment in the consonant coda (nasals being the only other coda segments). In other cases it has, in fact, been analysed as the only "consonant" in the coda. These distributional limitations have led to the present analysis.

Reference

Pei, Mario (1966): Glossary of Linguistic Terminology, New York: Doubleday & Company, Inc.



SOME OBSERVATIONS ON MUFAXXAMA THE "EMPHATIC" PHONEMES  
 IN ARABIC BY ROMAN JAKOBSON

Yousef El-Haleese, English Department, University of Jordan,  
 Amman, Jordan

These observations show that Jakobson's interpretations of the phenomena of emphasis in Arabic in terms of his distinctive feature theory have a number of shortcomings which affect his concept of binarism, the generalization validity of some of his features such as the opposition flat:plain, and the number of features used in his componential analysis of the "non-syllabic phonemes".

The opposition fortis:lenis which denotes voiced:voiceless overlaps with emphasis:non-emphasis.

Adopting his analysis we have to set up 48 consonantal phonemes instead of his 31, as all the 24 consonants found in the dialect under consideration can be emphatic and non-emphatic and they contrast in minimal pairs.

He assigns emphasis to consonants and neglects the vowels; and he just talks about two degrees of emphasis: emphatic and non-emphatic. The data upon which this paper is based makes it necessary to regard emphasis as a prosodic feature, the minimum domain of which is the syllable and the maximum can be a longer utterance. Five types of syllable can be distinguished with regard to emphasis:

1. Syllables emphasized all the way through
2. Syllables beginning with emphasis and ending with velarization
3. Syllables beginning with velarization and ending with emphasis
4. Syllables with velarization all the way through
5. Syllables with no emphasis.

Hence emphasis has to be regarded as a multivalued feature rather than binary, e.g. in da:ri "he is informed",

da:ri "he used to",  
dā:ri "my house".

The /d/ sound is non-emphatic, emphatic and velarized, respectively.

Consonants divide themselves into six classes according to their distribution initially and finally in these five types of syllables.

ZUR BESTIMMUNG VON PHONEMEN. EIN VORSCHLAG ZUR PRÄZISIERUNG  
DER 3. TRUBETZKOYSCHEN REGEL

Helmut Fasske, Akademie der Wissenschaften der DDR, Institut für  
sorbische Volksforschung, 86 Bautzen, Thälmann-Str. 6

Bei der Beurteilung der phonologischen Rolle von akustisch bzw. artikulatorisch verwandten Lauten setzt Trubetzkoy zwei Merkmale als Kriterien an: komplementäre Distribution und ausreichend differenzierende Invarianz.<sup>1</sup> Danach werden [x] und [ç] im Deutschen als kombinatorische Varianten eines Phonems, [h] und [ŋ] dagegen als Realisierungen zweier verschiedener Phoneme gewertet. Die Glieder beider Paare sind zwar komplementär distribuiert, den Lauten [h] und [ŋ] fehlt jedoch die ausreichend differenzierende Invarianz. Beide besitzen gemeinsam nur das Merkmal 'konsonantisch', das sie von anderen Konsonanten nicht ausreichend scheidet.

Bei einer strengen Anwendung der Trubetzkoy'schen Regel müssten auch ě und ǫ im Niedersorbischen als kombinatorische Varianten interpretiert werden. Sie haben eine komplementäre Distribution, ihr gemeinsames Merkmal (halbhohe Vokale) scheidet sie hinreichend von anderen Vokalphonemen. Dennoch werden ě und ǫ, bisher ohne Ausnahme, als Phoneme verstanden. Gibt es eine Begründung für eine solche phonologische Interpretation?

Eine Überprüfung aller entsprechender Fälle ergibt, dass alle als kombinatorische Varianten gewerteten 'verwandten' Laute nicht nur von allen übrigen Phonemen durch ein gemeinsames Merkmal differenziert sind, sondern dass sie sich voneinander durch ein solches Merkmal unterscheiden, das sich in keiner anderen phonologischen Opposition als distinktiv wiederholt, vorausgesetzt, dass diese Laute keine durch regressive Assimilation bewirkten Stellungsvarianten eines Phonems sind. In diesem Sinne wird ein Vorschlag zur Präzisierung der 3. Trubetzkoy'schen Regel unterbreitet.

---

(1) Trubetzkoy, N.S., Grundzüge der Phonologie, 3. Auflage,  
Göttingen 1962, 14

## SPELLING ERRORS AND LINGUISTIC CONSCIOUSNESS

Ivan and Peter Fónagy, C.N.R.S. (Paris), Department of Psychology, University College (London)

Through a sample of 10 000 spelling errors made by Hungarian children we attempted to make inferences about the linguistic consciousness of these children (Fónagy, I. and P., 1971).

Results

- (1) Spelling errors may reflect changes in functional relevance and might be indicative of phonetic change.
- (2) The omission of vowels, rare after 8 years, seems to result from a tendency to associate graphemes with syllables.
- (3) The correspondence between looser pronunciation and greater number of consonants omitted could be interpreted in terms of correspondence between the degree of phonetic distinctness and the degree of consciousness.
- (4) Consonant substitutions can be predicted by the phonetic assimilation laws. Errors showed no tendency on the part of the child to trace back superficial phonetic structures to possible underlying systematic phonemic ones.
- (5) Word separation errors reflect the child's consciousness of semantic units.

Reference

Fónagy, I. and P. (1971): "How to make use of spelling errors (in Hungarian)", *Nyelvör* 95, 70-89.

THE /-r/ SUFFIXATION AND THE PHONOLOGICAL STRUCTURE OF  
MANDARIN FINALS

Yi-Chin Fu, National Taiwan Normal University, Taipei, Taiwan,  
Republic of China

This paper attempts to show that the /-r/ suffixation in Mandarin Chinese is, in reality, a simple process involving only a single rule with a one-act operation - Terminal Truncation (Terminal = postvocalic glide or nasal) - rather than a "rather complicated" (Chao 1968, 46) phenomenon as has been claimed by many linguists. That is, the /-r/ suffixation in Mandarin always and only causes the loss of the Terminal segment of the underlying base forms of the finals. In order to make clear how our rule works, we present a new analysis of the underlying phonological forms of Mandarin finals. In essence, we have argued for a diphthong (or long) representation of the hitherto believed simple vowels /ɨ, e, a/ - i.e. /ɨh, eh, ah/, with /-h/ standing for a glide of the same quality as the preceding main vowel. We also propose a complex representation (Medial glide + diphthong) for the traditional simple vowels /i, u, y/, namely /jɨh, wiw, yɨh/. The mystery of why certain base forms should end up homophonous after /-r/ suffixation can be accounted for naturally, given our rule and our analysis of the base forms of the finals. Furthermore, with the availability of our rule and our base forms, all the /-r/ form listings like those of Hockett (1947), Chao (1968), and others, would be entirely dispensable, since all the /-r/ forms listed are now not only predictable but can be derived by rule in a simple way. The seeming complexity of the retroflex suffixation in Mandarin is brought to complete regularity.

References

- Chao, Y.R. (1968): A Grammar of Spoken Chinese, Berkeley and Los Angeles: University of California Press.
- Hockett, C.F. (1947): "Peiping phonology", Journal of the American Oriental Society 67, 253-267.

## PHONOLOGIE FONCTIONNELLE, PHONETIQUE EXPERIMENTALE ET DIALECTOLOGIE LUXEMBOURGEOISE - QUELQUES RESULTATS

Jean-Pierre Goudaillier, Laboratoire de Phonétique, U.E.R. de Linguistique Générale et Appliquée, Université René Descartes, Paris, France

L'utilisation des méthodes de la phonétique expérimentale, en relation avec les principes de la linguistique fonctionnelle, nous a permis de décrire en luxembourgeois un certain nombre de phénomènes, entre autres l'opposition de force articulatoire des occlusives /p t k/ et /b d g/ et l'adhérence phonique lors du passage d'une voyelle à une consonne subséquente.

L'étude de certains indices (sonorité, apparition des vibrations laryngiennes, durées, etc.) nous a amené à établir que le trait pertinent distinguant en Koïnè de Luxembourg-Ville /p t k/ de /b d g/ est la TENSION, la SONORITE et l'ASPIRATION n'étant que des traits distinctifs redondants (III occlusives). Nous avons pu déterminer la forme physique de l'unité restant en cas de neutralisation et celle des unités apparaissant dans les cas de liaison sonorisante.

L'examen d'une ligne d'intensité délivrée par un intensimètre nous a fourni la valeur des surfaces de chute d'intensité entre le(s) pic(s) d'une voyelle et le passage à la consonne subséquente, ce pour 375 séquences VC contenues dans des "mots". Nous avons ainsi le degré d'ADHERENCE PHONIQUE des unités de chacune des séquences; nous avons procédé au classement de celles-ci en fonction de leur degré d'adhérence ('contact') et avons établi des règles pouvant être résumées comme suit : dans une séquence VC, une diminution de la force vocalique entraîne une augmentation de l'adhérence phonique de la voyelle et de la consonne subséquente; une diminution de la force consonantique entraîne 1. une augmentation de l'adhérence de la consonne et de la voyelle précédente, si cette diminution de force consonantique n'est pas à l'origine d'une différence de force vocalique, 2. une diminution d'adhérence, si cette diminution de force consonantique provoque une augmentation de la force de la voyelle.

#### Références

Fischer-Jørgensen, E. & H.P. Jørgensen (1970) : "Close and loose contact ("Anschluß") with special reference to North-German", Annual Report of the Institute of Phonetics. University of Copenhagen, IV, p. 43-80.

Goudaillier J.-P., Phonologie fonctionnelle et phonétique expérimentale, Hamburg, Helmut Buske Verlag (à paraître).

## LA PALATALISATION EN FRANCO-QUEBECOIS

Hussein Habaili et Suzanne Richard, Université du Québec à Chicoutimi, Québec, Canada

Cette communication présente les résultats d'une recherche dont le but est d'élaborer un fragment de la composante phonologique du franco-québécois. Le problème qu'elle pose consiste à analyser à l'aide de règles phonologiques les plus générales possibles un processus phonologique du franco-québécois: la palatalisation.

L'intérêt de cette recherche réside dans le fait qu'elle essaie de dégager certaines particularités phonologiques du franco-québécois. Le cadre théorique et méthodologique dans lequel s'inscrit notre étude, celui du modèle générativiste abstrait, nous a permis grâce à son pouvoir descriptif et explicatif et ses principes de simplicité, d'économie et de généralité d'établir un ensemble de règles phonologiques qui tiennent compte de l'état actuel de ce processus et de son évolution.

Notre étude de phonétique expérimentale qui compte au-delà de trois cents palatogrammes réalisés auprès de sept informateurs nous a permis de faire le lien nécessaire entre phonétique expérimentale et phonologie. En effet, c'est grâce aux données objectives de cette analyse que nous avons pu déceler deux degrés de palatalisation, à savoir la palatalisation légère et la palatalisation forte. Tous deux formulables dans le cadre de la phonologie générative.

Ce dispositif formel nous a permis, après avoir dégagé certains universaux phonologiques relatifs à la palatalisation dans certaines langues naturelles - telles que le slave, le russe, le bulgare, l'italien, quelques langues amérindiennes et un parler arabe tunisien - de formuler une métarègle avec ses contraintes universelles qui rend compte du fonctionnement général de ce processus de palatalisation. Ce type de métarègle reflète d'une façon concrète les universaux phonologiques et constitue par le fait même l'expression formelle par excellence du caractère naturel des processus phonologiques.

ERSATZDEHNUNG UND PROSODIE IM BAIRISCHEN, SKANDINAVISCHEN  
UND ANDERSWO

Robert Hinderling und Anthony Rowley, Lehrstuhl für Deutsche  
Sprachwissenschaft, Universität Bayreuth

In unserem Vortrag geht es zunächst um die (oder eine) adäquate phonologische Interpretation der bairischen Silbenstrukturregel, die unter dem Namen "Pfalzisches Gesetz" bekannt geworden ist. Danach ist in diesem Dialekt Vokallänge immer mit Kürze oder Schwäche, umgekehrt Vokalkürze mit Länge oder Stärke des nachfolgenden Konsonanten kombiniert, ähnlich wie dies fürs Mittelschwedische gilt (Bannert 1976, S. 40). Einige bisherige Phonemisierungsvorschläge werden kritisiert und ein eigener, von Kufner (1957, 1961) ausgehender, näher begründet. Es ergibt sich, dass (oberflächenstrukturelle?) prosodische Unterschiede mit der Apokopierung in Zusammenhang stehen. Unser Lösungsversuch lässt sich darum auf anderes Sprachmaterial übertragen, bei dem ähnliche Apokopierungserscheinungen zu beobachten sind. Es wird ausser auf die Akzente des Schwedischen auf den Vestjysk stød, die Rheinische Schärfung und Erscheinungen des Ostpreussischen eingegangen.

Bibliographie

- Bannert, R. (1976): "Mittelbairische Phonologie auf akustischer und perzeptorischer Grundlage", Lund/Malmö.
- Hinderling, R. (1979): "Lenis und Fortis im Bairischen. Versuch einer morphophonemischen Interpretation" (im Druck).
- Kufner, H.L. (1957): "Zur Phonologie einer mittelbairischen Mundart", Zeitschrift für Mundartforschung 25, 175-184.
- Kufner, H.L. (1961): "Strukturelle Grammatik der Münchner Stadtmundart", München.
- Zehetner, L. (1978): "Die Mundart der Hallertau", Marburg.

## ON THE DYNAMICS OF /h/, EVIDENCE FROM FINNISH

Antti Iivonen, Department of Phonetics, University of Helsinki, Finland

Contrary to languages which have dropped the /h/, Finnish has established it as a new phoneme after the early proto Baltic-Finnic stage (cf. hän, iho, pihti, pihvi, inho).

The consonant /h/ is not involved in the quantity opposition (except hihhu and huhho), its participation in the boundary doubling rule is weaker than that of the other consonants (cf. istu+pa = [istup:a], but istu+han = [istuh:an] or more frequently [istuhan], all meaning 'do sit down'), and the stress signalling of h in /-Vh.C-/ structures shows some peculiarities.

Morphophonological alternations (kaksi/kahden, mies/miehen), some lexical relicts (löyhä/löysä 'loose', karkea/karhea 'rough'), diachronic changes, dialectal correspondences, omissions, assimilations, and substitutions in the first language acquisition reveal a number of similarities in the dynamics of /h/. Symptomatic interjections of the type huh-huh (sign for tiredness or fear) seem to show that the origin of /h/ is in an audible breathed sound which is phonemicized in marginal positions. /h/ is used also iconically as indication for frictional types of sound (huohottaa 'to puff', 'to pant'; suhina 'hum').

Diachronically h is the outcome from s, z, ʃ or the palatal tʃ or the velar plosive k. Final occurrences are diachronically avoided by metathesis (mureh > murhe 'sorrow').

Contrastively the syllable final cases of /h/ compared with German [ç] and [x] are interesting, especially the combinations of /h/ with the labial close vowels (Fi. nyhtää/Germ. nüchtern; Fi. puhti/Germ. Bucht) in which the Finnish h is produced with labial coarticulatory constriction (and corresponding "friction").

Phonetically: photo-electric glottograms show that the glottal opening for /h/ is greatest in /-h.C-/ structures (pihti, pihvi) where it is comparable with the first element in consonant clusters like /-k.s-/ or in double plosives (geminate) like /-k.k-/.

The usual voicing of h in /-h.C-/ structures (C being voiced) does not cause a homophony with /-VV-/ structures (cf. vihdan/viidan) and similar pairs can be distinguished in whispered speech. Whispered initial /h/ is also recognizable (cf. hosua/osua).



## DURATION OF ENGLISH SINGLE CONSONANTS AND CLUSTERS

A.A. Isengeldina, Academy of Sciences of the Kazakh Republic,  
Chair of Foreign Languages, Alma-Ata, USSR

1. Single consonants and clusters are shortest in intervocalic position. Next in duration are initial consonants, and longest are final consonants and clusters.

2. In initial two member clusters the first component is longer than the second, except for /l/ and /r/, which tend to be longer than the first consonant in the cluster.

In intervocalic two member clusters the duration of the components is dependent on the word accent, a consonant in a stressed syllable always being the longest one.

In final clusters the last consonant is the longest, in clusters sonants followed by a voiced consonant are approximately twice as long as sonants followed by voiceless consonants.

3. In initial three member clusters the duration of the consonants decreases from the first to the third component according to a ratio 3:2:1.

In intervocalic clusters the second component (an obstruent) is the shortest; the duration of the first and third components depends on the word accent, the consonant belonging to the stressed syllable being the longest.

In final three member clusters the second consonant is the shortest and the last one is the longest. If the final cluster consists of voiced consonants, it is at least twice as long as the corresponding cluster with voiceless consonants.

## DIRECTIONALITY AND EXCEPTION-FEATURES IN GENERATIVE PHONOLOGY

Richard D. Janda, Department of Linguistics, University of California, Los Angeles, U.S.A.

Recently, phonologists have been increasing their efforts to endow (parts of) grammars with a directionality (uni-, bi-, or multi-; cf. Eliasson (1978)); concomitantly, generative phonologists of widely varying persuasions all seem to be intensifying their attempts to eliminate exception-features (cf. Chomsky and Halle (1968, passim)) from their descriptions of languages. The present paper, however, adduces evidence which, we believe, shows both of these goals to be fundamentally misguided and worthy of speedy abandonment--at least in a competence-oriented model of phonology.

In demonstrating this, we will start with a more general perspective and show that: (1) so-called "Standard" Generative Phonology is essentially non-directional, despite the claims of Eliasson and the occasional practice of Chomsky and Halle themselves, and (2) the elimination of exception-features is undesirable because it removes a theoretical device that plays the indispensable role of distinguishing what is exceptional and ad hoc from what is not. We will then conclude with a more concrete example which shows that the newly proposed theory of "Upside-Down" Phonology--which attempts, essentially, to achieve both of the above-mentioned goals simultaneously--is, in fact, unworkable in principle: reversing what is commonly taken to be the directionality of SGP (viz., from underlying to surface representation) does not allow one to dispense with exception-features, in either synchronic or diachronic phonology. The fact that both directions of derivation require the same kind of mechanisms for handling exceptions thus leads inescapably to the conclusion that generative phonology actually is non-directional--a significant finding.

#### References

- Chomsky, N. and M. Halle (1968): The sound pattern of English, New York: Harper and Row.
- Eliasson, S. (1978): "Directionality in lexical signs and grammars: remarks on the emergence of a theoretical concept", SL 32, 50-62.
- Leben, W. R. and Robinson, O. W. (1977): "'Upside-down' phonology", Lg. 53, 1-20.

## ON THE COMPONENTS OF THE DISTINCTIVE FEATURES OF PHONEMES

Z.N. Japaridze, Institute of Linguistics of the Academy of Sciences of the Georgian SSR, Tbilisi, USSR

It may be assumed that the distinctive features of phonemes (at any rate, some of them) dissolve into smaller linguistic units. The fact that these units are variously "assessed" in different languages may be taken as an index of their linguistic nature.

Thus, the feature "continuant - interrupted" can be split into at least two components: one connected with the duration of noise and the other with the rate of noise intensification. At a constant duration of noise, depending on the rate of noise intensification, sounds can be perceived as continuant or interrupted. It is feasible to synthesize a sound that is perceived by a Georgian either as spirant s or as affricate c, depending on the manner of reproduction: from beginning to end or from end to beginning.

The change of a spirant into an affricate is attained by removal of both the beginning and the end. In the latter case only the duration changes, whereas in the former both components are altered. The values of components compensate for each other. The effect of compensation depends here not only on the value of their parameters, but also on their relative weights in any given language. Obviously, the relative weights of noise intensification should be different in languages where this characteristic may differentiate sounds of the type s, c, t, and in languages where sounds of the type c (affricate) are absent. The weights of the noise duration component in languages where long and short phonemes may or may not be differentiated by noise duration, should also be different.

A parameter of distinctive feature with several components can be represented as a sum of the values of these components. The value of each component has its own coefficient. The latter reflects the component's relative weight, i.e. its power of compensation in a given language. In contrast to this, different distinctive features cannot compensate for each other.

## FROM PHONEME TO SPEECH SOUND: CONSTRAINTS IN THE LINKAGE OF VARIABILITY IN ALLOPHONY

Ilona Kassai, Institute of Linguistics, Budapest, Hungary

In the course of analysing Hungarian consonant clusters, several correlations were discovered, which, should they prove correct, can contribute to our understanding of how competence passes into performance. This presentation will set forth some of the most important of these correlations.

Facts determining the realizations of phonemes can be linguistic or non-linguistic, the former either segmental or supra-segmental. The speech sound realizations of phonemes are fundamentally determined by the regularities of the syntagmatic and paradigmatic axes. Within the frame given by the requirements of the paradigm (maintenance of self-identity), the requirements of the syntagma (identity with other [neighboring] phonemes) operate automatically given the absence of other - especially suprasegmental and extra-linguistic factors. The status of the phonemes within the system (integrity, to use Martinet's term) bears the following relation to the speech sound: The more distinctive features mark a phoneme, the more stable, i.e. identical, are its realizations. The paucity of features is equal to the increasing influence of the syntagmatic axis. The following corollaries may be drawn from this observation: 1. The greatest influence can be perceived between those elements that differ only by one feature, though the physiological and perceptual nature of the feature determines its exact extent. 2. The nature of the speech sound is in close agreement with the nature of the producing organ. The more mobile the organ, and the fewer organs required, the more the sound in question adjusts itself to its surroundings.

This syntagmatic principle operates without constraints in the case of two features, voicing and length, that embrace the entire consonant system. The effects, however, operate unequally, depending on the marked or unmarked nature of the element in question. Thus, in the case of both voicing and length, the neutralization prefers the lesser marked member of the opposition.

## DIE MUTATION a/o IN NORDRUSSISCHEN MUNDARTEN

Evgenij P. Kirov, Philologische Fakultät der Kazañer Universität,  
Kazañ, USSR

Die vielen nordrussischen Mundarten eigene phonetische Entwicklung a > o in vortonigen Silben ist teils durch Koartikulation mit Lippenlauten und Hinterzungenlauten [bogáŝ], teils durch Assimilation zu den betonten Vokalen [o] und [u] [norót] verursacht.

Die phonetische Entwicklung a > o führt in einigen Mundarten zu einer Vereinfachung des Vokalismus. Es handelt sich um die Verstärkung der Relevanz der phonologischen Eigenschaft Rundung, und eine Reduktion der Relevanz von Artikulationsstelle und Öffnungsgrad. In diesen beiden Dimensionen ist a das schwächste Element.

Der ungespannte und offene vortonige Vokal [a] wird unter diesen Bedingungen immer gerundet. Dazu trägt die Nichtrelevanz der Rundung in vortonigen Silben mit dem Phonem /a/ und mit dem Hyperphonem /a/o/ bei. So entsteht eine phonologische Änderung a > o [sož'éŋ]. Sie kann in der Terminologie von R. Jakobson (1931) als eine Mutation a/o bezeichnet werden.

Die phonologische Entwicklung a > o umfasst einen grossen Teil des Wortschatzes. Eine statistische Analyse der Umgangssprache (300 Wörter mit vortonigem [a]) gab folgende Ergebnisse: Idiolekt I. 94%, Id. II 92.4% und Id. III 93% Änderung a > o.

Literaturhinweis

Jakobson, R. (1931): "Prinzipien der historischen Phonologie", Travaux du Cercle linguistique de Prague IV, 247-267.

## PHONETIC EUPHEMISMS AND PHONOLOGICAL DISTANCE

John M. Lipski, Dept. of Romance Languages, Michigan State University, East Lansing, Michigan, 48824, U.S.A.

In the study of phonology, a number of models have arisen to depict the operation and internal structure of the phonological component, each model based on a differing set of fundamental postulates and bearing differing relations to the realm of empirical evidence. In particular, one may start either from a purely axiomatic basis or from an uncorrelated set of empirical data, with the majority of studies coming from intermediate points on this methodological continuum. Once values for particular distinctive features have been determined, it is possible to establish matrices of phonological 'distance' based on feature values. In this study, an additional method is suggested by means of which one may hope to gain insight into the workings of distinctive feature systems. The phenomenon in question is the phonetic modification of socially proscribed words to form acceptable euphemisms. Obviously, constraints exist which allow some types of modification and disallow others. One particular case, from Spanish, is analyzed in detail, and after examining purely phonetic modifications (as opposed to lexical replacement by already existing forms), it is concluded that such constraints may indeed be discovered by further study of existing and potential euphemistic deformations. An ongoing research project is revealing both anticipated and unexpected specifications of phonological distance, suggesting the need to reexamine the notions of distance and the irrelevance of individual phonological specifications in determination of phonological systems.

POUR UNE EXPLICATION "NATURELLE" (ACOUSTICO-ARTICULATOIRE) DE  
LA MUTATION u > y

F. Lonchamp et F. Carton, Institut de Phonétique, Université  
de Nancy II, 54015 Nancy, France

L'explication du passage [u > y] a fait l'objet de plusieurs hypothèses (chaîne de pression, etc...) qui ne nous paraissent pas convaincantes. Dans la ligne des recherches actuelles en phonologie "naturelle" (cf. Ohala 1974), une simulation sur ordinateur a montré que l'antériorisation a pour objet de rétablir l'opposition [o - u] mise en péril par la délabialisation qui provoque une élévation de F1 et F2. Plusieurs indices confirment notre thèse: le japonais associe antériorisation et délabialisation pour [w]; la faible labialisation de l'australien provoque actuellement cette mutation. On note une résistance de [u] en contexte consonantique labial (franco-provençal, ancien français) ou en l'absence d'affaiblissement en [v] de [w] et [p, b] (wallon). L'alsacien a connu simultanément les mutations [y > i] et [u > y]. [y < u] aboutit souvent rapidement à [i] (grec ancien, etc...), ce qui suggère plutôt [u > i], puis relabialisation éventuelle en [y], avec enfin [o > u].

Références

- Lonchamp, F. (1978): Recherches sur les indices perceptifs des voyelles orales et nasales. Thèse de III<sup>e</sup> Cycle - Université de Nancy II.
- Ohala, J. (1974): "Phonetic explanations in phonology", Chicago Linguistic Society, 251-274.

## THE NATURE OF LINGUAL CONSONANTS

E. N. A. Mensah, Department of Linguistics, University of Ghana,  
Legon, Ghana

One of the deficiencies of the standard Distinctive Feature System is the absence of a feature for alveolars, palatals, and velars: in fact, in the Chomsky and Halle system alveolars and velars are maximally opposed to each other. In the paper it is shown that this maximum opposition is unnatural as these consonants often undergo the same phonological processes. Since these consonants are all actively produced with the tongue, it is proposed that they be classified by the feature lingual. It is also shown that /h/ is a lingual and not a "glottal fricative", as often claimed in phonetics.

Of all the phonological processes that go to define linguals as a natural class, palatalisation is perhaps the commonest: most linguals become complete palatals, whereas non-linguals may become only partially palatalised. Examples are drawn from a number of West African languages to support this claim. For instance, in the Akan language, k, g, w, kw, gw, h become c, ɟ, ɥ, cw, ɟw, ɥ, respectively, and the palatalisation of labials is only restricted to the addition of a secondary palatal /j/ to the primary articulation. But more particularly the complete change of /h/ to ɥ indicates that the former is inherently a lingual.

In most West African languages there is phonological evidence to show that /h/ is in addition a velar as it regularly alternates with /k/, especially in consonant mutation.



## ASPIRATES AND MAHĀ-PRĀNA IN SINDHI

Paroo Nihalani, Central Institute of English and Foreign Languages, Hyderabad, India

Chomsky and Halle (1968) have invoked the feature [+ heightened subglottal pressure] to characterise the contrast between aspirated & unaspirated and murmured & non-murmured sounds. They claim that [ph], [bh] etc. are produced with heightened subglottal pressure, and their unaspirated counterparts without it. This claim is controverted by experimental data from Sindhi, which shows that although [bh, dh, ḍh, gh] are produced with higher subglottal pressure than their non-murmured counterparts, the palatal sound [ɟh] has less subglottal pressure than its non-murmured counterpart [ɟ].

It is therefore proposed here to introduce the feature [+ increased airflow rate] i.e. 'mahā-prāna' to replace Chomsky and Halle's feature [+ heightened subglottal pressure] which does not seem to be phonetically well-motivated. Instrumental findings clearly indicate that both 'voiceless aspirated' and 'murmured' sounds differ from the 'voiceless unaspirated' and 'voiced' sounds by having higher airflow rate, which concurs with the phonetic labelling of the distinction between aspirate and non-aspirate, namely 'mahā-prāṇa' and 'alpa-prāṇa' as suggested by the ancient Hindu grammarians.

## EVALUATION OF THE CORRELATIVE OPPOSITION OF "SOFTNESS"

Jiřina Novotná-Hurková, Czechoslovak Academy of Sciences,  
Prague, Czechoslovakia

The paper is based on a comparison of the sound material of three West Slavic languages: Czech, Polish and Upper Lusatian Sorbian.

The basic difference between these languages is on the one hand in the very quality of the studied correlative opposition and, on the other hand, its quantitative utilization. In the majority of softness pairs in Polish and in Upper Lusatian Sorbian (as well as in other languages, in whose phonological systems the correlative opposition of softness and hardness is firmly anchored), we are dealing with a phonetic difference between palatalized and non-palatalized speech sounds, while in Czech softness pairs we have a phonetic difference between a palatal and non-palatal speech sounds.

We have attempted, on the basis of phonetic and phonological analysis, to define the correlative softness opposition in Western Slavic languages also from the point of view of the central phonological system and of the periphery of the languages studied.

References

- Borovičková, B. and V. Maláč (1967): The Spectral Analysis of Czech Sound Combinations, Praha, Academia.
- Kučera, H. (1961): The Phonology of Czech, s'Gravenhage.
- Romportl, M. (1966): "Zentrum und Peripherie im phonologischen System", Travaux linguistiques de Prague 2, 103-110.
- Vachek, J. (1968): Dynamika fonologickeho systému současné spisovne češtiny, Praha, Academia.
- Wierzchowska, B. (1967): Opis fonetyczny języka polskiego, Warszawa.

## THE PHONOLOGICAL FEATURES OF HINDI STOPS

Manjari Ohala, San José State University, San José, Calif., USA

The phonological features chosen to represent segments should a) reflect natural classes among the segments, and b) be non-arbitrary, i.e., empirically verifiable. The features proposed by Chomsky and Halle (1968) and Halle and Stevens (1971) for the series of obstruents found in Hindi more or less meet the first requirement; do they meet the second?

Using a plethysmograph to measure lung volume and a pneumotachograph to sample oral air flow in the speech of one Hindi speaker, I sought to verify Chomsky and Halle's claim that aspirated stops are differentiated from others by the feature of 'heightened subglottal air pressure', which logically implies an active pulmonic gesture. Contrary to their claim I found no evidence of any active pulmonic involvement in the production of these segments. Rather, variations in the rate of lung volume decrement during all obstruents, aspirated or not, can be attributed to passive reaction to variations in the air flow escaping from the lungs due to variations in glottal and supraglottal impedance. The finding of high rate of flow upon release of [p<sup>h</sup>] and [b<sup>h</sup>], however is compatible with Halle and Stevens' claim that the vocal cords are abducted for these stops.

To test Halle and Stevens' claim that both [b] and [b<sup>h</sup>] are produced with slack vocal cords, fundamental frequency (F<sub>0</sub>) on the vowels flanking medial stops and sonorants was sampled and averaged over 10 tokens of each consonant type. F<sub>0</sub> was significantly lower on the vowel following [b<sup>h</sup>], but there was no appreciable perturbation of the F<sub>0</sub> on vowels near the other stops. Thus Halle and Stevens are correct in their contention that vocal cord tension is used distinctively in Hindi stops but wrong in assuming that [b] and [b<sup>h</sup>] have the same tension. Revised feature specifications of Hindi stops that meet requirements (a) and (b) above will be presented.

#### References

- Chomsky, N. and M. Halle (1968): The sound pattern of English, New York: Harper & Row.
- Halle, M. and K.N. Stevens (1971): "A note on laryngeal features", Q. Progress Report, Research Lab. of Electronics, MIT, 101, 198-213.

## SPRACHLAUT, SCHRIFTZEICHEN UND PHONEMWANDEL

Herbert Penzl, Department of German, University of California, Berkeley, California 94720, USA

Auch die Beziehung zwischen akustischen (Lauten) und schriftlichen Sprachzeichen gehört zur Phonologie. Schreibungen ergeben synchronisch und diachronisch wichtiges Material, dessen Erfassung, wie hier behandelt werden soll, durch die Aufstellung einer Typologie gefördert werden kann.

An Abweichungen von der Norm eines alphabetischen Schreibsystems sind mechanische Verschreibungen, auch systembedingte Varianten (deutsch Teater für Theater) linguistisch selten relevant, wohl aber sind es nichthochsprachliche Nebenformen (engl. ruther 'rather'), Schnellformen (unsre für unsere) und alle Arten von Ausspracheschreibungen (engl. iland für island, feudal für futile; deutsch Proplem 'Problem' usw.).

Eine rein graphisch orientierte Beschreibung nach der Änderung der Schriftzeichen oder eine "pragmatische" nach dem angenommenen Sprechakt, bzw. dem Vorkommen in den einzelnen Textsorten ergäbe keine systematische Typologie. Manche Schreibungswandlungen sind natürlich ohne jede phonologische Motivierung. Ist sie aber vorhanden, lässt sich eine Typologie zwar nicht nach den generativistischen Regeländerungen, wohl aber nach den strukturalistischen Haupttypen des Phonemwandels (wie bei Jakobson, Hoenigswald, Jones, Martinet, Penzl, Moulton) aufstellen. Weglassen, Einschub, Umstellung, Angleichung von Buchstaben entsprechen den phonotaktischen Änderungen. Schreibungsschwund und Graphisierung bezeichnen Phonemschwund, Schreibungsersatz die Phonemverschiebung, Schreibungsüberschneidung, Schreibungszusammenfall und Schreibungsumkehrung den Phonemzusammenfall, Schreibungsspaltung die Phonemspaltung. Schreibungsumwertung (mittelengl. i in time später für /aI/, usw.) deutet auf Verkettung von Phonemwandlungen ("Schub", "Sog").

Das diachronische Beweismaterial weist darauf hin, dass die Beziehung Schriftzeichen/Sprachlaut innerhalb eines historisch gegebenen alphabetischen Schreibsystems nicht als willkürlich ("arbiträr") angesehen werden kann.

Litteraturhinweise

Penzl, Herbert (1972): Methoden der germanischen Linguistik, Tübingen: Niemeyer.

Penzl, Herbert (1969): Geschichtliche deutsche Lautlehre, München: Hueber.

## NON-SYLLABIC PHONOLOGY

Herbert Pilch, Albert-Ludwigs-Universität, Freiburg im Breisgau

Problem: Establish the phoneme paradigms of English beyond the initial consonants (including clusters), vowels (including polyphthongs) and final consonants of monosyllables. There are very obvious limitations, for instance no vocoid other than [ə] before enclitic ("unstressed") /ʒ ŋ pɪ/. The phonological structure involved can thus not be interpreted as a sequence of monosyllables with the same initial consonants, vowels and final consonants as the monosyllable.

Solution: Replace the syllable, as an analytical notion, by the phonemic Shape Type. The enclitic shape (such as knowledgeably, platypuses) has seven slots, each with its specific inventory. The analysis is based on English as spontaneously spoken, not as codified in Pronouncing Dictionaries.

## DUTCH MARGINAL PHONEMES AND THE ADAPTATION OF ENGLISH LOANS

Jan Posthumus, Department of English, University of Groningen,  
The Netherlands

A study of the adaptation of English loans highlights the existence of a fairly extensive, already well-established set of loan phonemes, readily drawn upon by the educated Dutch speaker for the pronunciation of foreign loans.

Qualifying for membership of this set are nine vowels and three consonants, most of them related in a fairly simple way to members of the primary system. In current accounts of Dutch phonology these marginal phonemes only receive scant attention. This paper will examine their status within the Dutchman's total inventory by seeking answers to the following questions: 1. In approximately how many words does the phoneme occur? 2. Are there common words among them used by all strata of the population? 3. To what extent do we find alternative pronunciations in which the loan phoneme is replaced by a member of the primary system?

It will be further pointed out how the influx of the more recent English loans is affecting the status of each loan phoneme. Attention is also drawn to certain areas of indecision in the adaptation of the English phonemes.

The following conclusions are drawn: Of the phonemes in question, /ɛ:/ and /ɜ,ʒ/ are irreplaceable members of the Dutch speaker's system; /ɔ:/ and /i:/ are well-established in certain words, but have alternatives in others; the same holds for /u:/, which, however, occurs in rather fewer words, now mostly English; /g/ is, on the whole, easily replaceable by /ɣ/ in older, mostly French loans, though not yet in English loans; /æ:/ and /ɤ:/ are so rare that they are little more than curiosities; lastly, the use of nasal vowels in French loans marks the well-educated speaker who knows his French: replacement by oral vowel plus nasal is practically always possible.

## VOWEL HARMONY IN NATURAL GENERATIVE PHONOLOGY

Catherine O. Ringen, University of Iowa, Iowa City, Iowa, USA

According to Hooper (1976, 1977) a central constraint in the theory of natural generative phonology (NGP) is the true generalization condition (TGC) which states that all rules must make true generalizations about surface representations. Thus, according to the TGC, a rule such as

$$(1) V \rightarrow [\alpha \text{ back}] / \left[ \begin{array}{c} V \\ \alpha \text{ back} \end{array} \right] C_0 \text{ ---}$$

is not a possible rule in language L if there are surface representations which contain vowels differing in backness (e.g. tati). Such a rule (if it is a rule at all) must be formulated in NGP as a morphophonological rule (MP-rule) which makes reference to morphological, syntactic, or lexical features.

The purpose of this paper is to consider how vowel harmony rules such as (1) must be (re)formulated in NGP to meet the TGC. It is argued that the analyses required by NGP for vowel harmony in languages such as Turkish, Hungarian, Finnish, and many West African languages (e.g. Igbo) are incorrect and do not reflect the generalizations which speakers of these languages have made. Specifically, it is shown that according to NGP vowel harmony in a language like Turkish or Hungarian must be treated as a non-productive, suppletive alternation. There is, however, strong evidence from language change, from acquisition, and from the treatment of loanwords that vowel harmony in these languages is a productive phonological rule and that the NGP analysis is incorrect. It is concluded that vowel harmony systems provide strong evidence against the TGC and thus against the theory of NGP.

#### References

Hooper, J. (1976) An Introduction to Natural Generative Phonology, New York: Academic Press.

(1977) "Substantive principles in Natural Generative Phonology," paper presented at the Conference on the Differentiation of Current Phonological Theories; to appear in Current Phonological Theories, D. Dinnsen, ed., Bloomington, Indiana: Indiana University Press.

## LINGUISTIC INTUITION AND PHONOLOGICAL DATA

Jon D. Ringen, Indiana University, South Bend, IN, USA

In this paper, it is shown that intuitive linguistic judgments are in fact used as a source of phonological data and that these judgments are no different in kind from those used in syntax and semantics. It is argued that given the goal (endorsed by all generative linguists) of characterizing rules which actually govern speech (and which speakers tacitly know) intuitive judgments are indispensable (even in principle) as data for evaluating phonological theories. These considerations suggest that, despite linguists' apparent lack of concern with understanding the use of intuitive data in phonology, recent discussions of the nature of linguistic intuition, of the scientific legitimacy of using linguistic intuition as data, of the epistemic status of theories in whose evaluation intuitive judgments play a significant part, and of methodologies appropriate for assessing the relative reliability of conflicting intuitive linguistic judgments (and of speech and judgments which are in conflict) are as relevant to the conceptual and methodological foundations of generative phonology as they are to generative syntax and semantics.



## ZUR DISTINKTIVEN OPPOSITION "PALATAL - NICHTPALATAL"

Milan Romportl, Institut für Phonetik, Karlsuniversität Prag,  
Tschechoslowakei

Unsere Modifizierung der Theorie der distinktiven Merkmale<sup>1</sup> wird zur Analyse der konsonantischen distinktiven Opposition "palatal-nichtpalatal" bzw. "weich-hart" angewendet, die in einigen slavischen, ausnahmsweise auch nichtslavischen Sprachen zur Geltung kommt. Es interessiert uns dermassen weder das Problem, ob man in einzelnen Sprachen diese Opposition für eine Korrelation halten soll oder nicht, noch die Frage der Proportionalität dieses Gegensatzes. Es wird vor allem das Problem erörtert, durch welche akustische Mittel diese Opposition realisiert wird.

Auch dieses Merkmal wird nicht nur durch eine einzige akustische Eigenschaft, sondern durch eine ganze Menge solcher Eigenschaften charakterisiert. Die "weichen" Glieder werden durch eine grössere Dauer des Explosionsgeräusches - QEx -, ein höheres Zentrum dieses Geräusches - CEx -, eine grössere Dauer des konsonantischen Segments - QCons -, eine ausgeprägte Transientphase des  $F_2$  -  $TF_2$  -, sowie auch des  $F_1$  -  $TF_1$  -, bzw. auch  $TF_3$  usw. gekennzeichnet.

In unserer Transkription,<sup>2</sup> wo auch die Hierarchie der Eigenschaften ausgedrückt wird, kann man die Struktur dieses Merkmals folgendermassen darstellen:

$$\left\{ \begin{array}{c} TF_2 \\ CEx - QEx - TF_1 \\ (QCons) - (TF_3) - \dots \end{array} \right\}$$

- 
- (1) Vgl. Romportl, M.: Zvukový rozbor ruštiny, Prague 1962; ds.: Studies in Phonetics, Prague-The Hague 1973; ds.: "Zur Struktur der phonologisch distinktiven Merkmale und der distinktiven Oppositionen", in Bereiche der Slavistik-Festschrift Josip Hamm, Wien 1975, 253-260; ds.: "Neueres über die akustischen Korrelate der distinktiven Merkmale", Phonologica 1976, Innsbruck 1977, 239-242.
- (2) Z.B. Romportl, M.: Studies in Phonetics, S. 17ff.

## VOYELLE, SEMI-VOYELLE ET CONSONNE

A. Rosetti, Bucarest

Le but de notre communication est de confirmer la classification des voyelles, semi-voyelles et consonnes que nous avons proposée en 1942 et dans les années suivantes: du point de vue fonctionnel, en phonologie, les semi-voyelles i et u jouent le rôle de consonnes.

Dieter Meinert et Eberhardt Richter, qui ont examiné récemment ce problème, sont d'un avis contraire: du point de vue fonctionnel, y et w ne jouent pas le rôle de consonnes.

L'argumentation des deux auteurs repose sur l'analyse instrumentale des sons parlés, qui ne peut pas fournir des unités et des oppositions phonologiques. Leur erreur est d'avoir appliqué aux phonèmes, situés à un autre niveau de l'analyse, des résultats concernant les sons parlés.

## DISTINCTIVE FEATURE CONSTRAINTS ON PHONEME ERRORS OF DIFFERENT TYPES

Stefanie Shattuck-Hufnagel, Cornell University, Ithaca, NY and  
Dennis H. Klatt, MIT, Cambridge, MA.

The substitution of one phoneme for another in spontaneous speech errors can take the form of (1) the Exchange of two target segments (as in "top shalk" for "shoptalk") or (2) the Substitution of an intrusion segment for a target. Substitutions may be Anticipatory (as in "Rynn rang" for "Lynn rang"), Perservatory (as in "knee neep" for "knee deep") or No-source (as in "Winken, Blinken & Mod" for "Nod").

Most analyses have combined all error types into one consonantal confusion matrix. Yet, various production models in the literature make different predictions about feature constraints on different error types. To test these predictions, a corpus of 820 consonantal errors was divided into separate matrices by error type, and each matrix analyzed by the method of Klatt (1968). For the three dimensions of voicing (2 values), manner (6 values) and place (6 values), the exchange and anticipatory substitution matrices are indistinguishable. In contrast, perseveratory substitutions preserve the place feature and no-source substitutions preserve the manner feature significantly more often. Analyses of a larger corpus, using a number of alternative feature systems, are underway.

Earlier studies have also shown that intrusion and target segments share distinctive features more often than would be predicted by chance. This has been interpreted as support for the claim that distinctive features, rather than phones, move and exchange in errors; details of the exchange error data permit us to refute this claim (Shattuck-Hufnagel & Klatt, 1979).

#### References

- Klatt, D.H. (1968), "Structure of confusions in short-term memory between English consonants", JASA 44, 401-407.
- Shattuck-Hufnagel, S.R. and Klatt, D.H. (1979), "The Limited Use of Distinctive Features and Markedness in Speech Production: Evidence from Speech Error Data", J. Verb. Learn. Verb. Behav. (in Press).

## THE RISE OF THE NEW DIPHTHONGS IN MIDDLE ENGLISH

Valeriya Sirokhvatova, Karelian Pedagogical Institute,  
Petrozavodsk City, USSR

It is generally assumed that the biphonemic combinations consisting of stressed vowels and vocalized spirants gradually turned into monophonemic diphthongs in the course of the change of the opposition of quantity into that of contact (Vachek, 1959).

Another version of this phonological phenomenon can be given. The point is that the rise of the new ME diphthongs can be regarded not as a gradual process but a qualitative leap. All the existing biphonemic combinations turned into monophonemes just after the loss of the final unstressed /e/, the forms in question thus having ceased to be divisible into syllables and morphemes. At the first stage of this change the initial elements of the diphthongs could be phonetically identified with the isolated vowels they had descended from.

At this time length and degree of aperture ceased being relevant because of the decay of the opposition of quantity. In connection with it all members of the vowel system (including the new diphthongs) underwent the corresponding changes but since the diphthongs were now functionally independent phonemes, the results of these changes in the diphthongs were unlike those in the isolated vowels.

Reference

Vachek, J. (1959): "Notes on the Quantitative Correlation of Vowels in the Phonematic Development of English", Mélanges F. Mossé in memoriam, Paris, 444-456.

## ON MINIMIZING FEATURE SPECIFICATIONS OF PHONEMES

William J. Sullivan, Program in Linguistics, University of Florida, Gainesville, Florida 32611 USA

This study tells what structurally justifiable orders of feature specification can be used to eliminate redundant features for the phonemes of Russian and what is the minimum number of features necessary for the unique specification of each phoneme.

Background and summary

Jakobson, Cherry, and Halle 1953 specify the 42 phonemes of Russian fully, using eleven binary distinctive features. They show that altering the order of feature specification for different phonemes reduces the average number of features specified per phoneme to 6.5. But they do not justify these different orderings on any structural basis. Next, they show how the 6.5 features/phoneme can be reduced to 3.05 for triphonic groupings by considering sequential constraints. Extensions of the methodology are indicated.

This study uses a generalized description of the Russian syllable, which is probably the limiting phonotactic case. Syllable structure determines the order of specification of features and specifies different sets of features for different classes of phonemes. Each phoneme is directly related to only one feature. All other features pertinent to a given phoneme are supplied redundantly by the phonotactics as a function of the way that phoneme is related to the syllable.

Conclusions

The tactic structure of the Russian syllable predicts the specification of redundant features and identifies each phoneme uniquely by directly relating it to only one feature. This seems to be independent of the feature system chosen.

Reference

Jakobson, R., E.C. Cherry, and M. Halle (1953): "Toward the logical description of languages in their phonemic aspect", Lg. 29, 34-46.

## DEFINING THE PHONEME: PHENOMENOLOGICAL ASPECTS

Tamás Szende, Institute for Linguistics, Budapest, Hungary

The concept of the phoneme as used by the Prague School and later structuralists does have a post-Chomskyan future, provided that it is made explicit in phenomenological terms. Phenomenological definitions of the phoneme center on its 'manner of existence'. These definitions form an implicative sequence, in that the *n*th implies the *n* - 1st. Some of the more basic ones are suggested herein.

(1) Phonemes are existent in the sense of being given and being independent of any given individual's recognition of their existence.

(2) The prime attribute of the phoneme is the constancy of its identity at all given points in a communicative event.

(3) Phonemes exist in language as abstracta while allophones represent a subordinate class of their concrete realizations.

(4) Phonemes are to be considered abstract elements because they account for concrete events.

(5) The 'manner of existence' of phonemes is to be found in their linguistic relevance based on both the constancy of their identity and their relevant functioning.

(6) Since phonemes form a linkage to biological events, they are of a symbolic nature.

(7) Phonemes are existent in individual units in the communicative event, which is to say that the existence-relation of the phonemes and the corresponding speech sounds is one of mutual dependence.

## DURATION DIFFERENCES AS A CUE FOR CONSONANT GRADATION IN LAPPISH

Brit Ulseth, Department of Linguistics, University of Trondheim,  
N-7000 Trondheim, Norway

Introduction

In Lappish as in e.g. Finnish there is a systematic change of the stem consonant in words which belong to the same inflection category. This change is found throughout the vocabulary. It functions as a marker of case for nouns and as a marker of person and number for verbs.

The aim of the present investigation has been to detect the possible role of duration differences as a cue for consonant gradation within a limited area of Lappish, viz. the Jukkasjärvi dialect spoken in the Kiruna district in Swedish Lappland.

Material and method

Recordings were made of 10 adult male speakers who read a word list of 28 words. The words were of the type  $(C_1)V_1C_2V_2$ , where  $C_2$  is a dental/alveolar stop or fricative. All the words were said in the same frame sentence. Sonagrams were made, and the duration of the segments  $V_1$ ,  $C_2$ , and  $V_2$  was measured.

Results

The results show that there is a close and regular relation between the duration of the stem consonant/parts of the stem consonant of the strong and the weak grade.

Phonetically, the results may be regarded as making up two groups. In one group the duration of  $C_2$  is significantly longer in the strong grade than in the weak grade. The other group may be divided into two sub-groups: (1) the pre-aspiration part of the consonant in relation to the closure part of the consonant is longer in the strong grade than in the weak grade, and (2) the voiced part of the closure in relation to the unvoiced part of the closure is longer in the strong grade than in the weak grade. The results from a listening test seem to confirm these results.

Phonemically, the results of the material investigated seem to suggest the induction of the following tentative generative rule in Lappish: There is an element in the stem consonant whose duration in relation to the rest of the consonant is a predominant factor both productively and to some extent perceptively. This part of the stem consonant is of vital importance to divide the strong grade from the weak grade.

SOME LOGICAL CONTRADICTIONS IN THE THEORY OF THE PHONEME  
REGARDED AS A BUNDLE OF DISTINCTIVE FEATURES

Galina Voronkova, Institute of Scandinavian Languages, University  
of Leningrad, USSR

This paper attempts to show the inadequacy of the "bundle theory" of the phoneme.

Subjects

The theory of the phoneme regarded as a bundle of distinctive features deduced from the oppositions among phonemes, which was advanced first by Prague phonologists, has become widespread and used also in the phonology of the Norwegian language. There is, however, a certain contradiction between the theoretical premises according to which a phoneme is a term of all oppositions and the methods of determining its phonological content. The distinctive features of phonemes are usually determined not on the level of the whole system but on the level of a subsystem. Thus the so-called phonological (linguistic) content of a phoneme based on the oppositions into which the phoneme in question enters consequently is only part of its phonological content. The method of determining the distinctive features of phonemes having no correlatives has some logical errors.

Conclusion

Both the assumption that the phoneme is a bundle of distinctive features and that distinctive features reflect the linguistic nature of the phoneme can be called in question.



## WORD BOUNDARIES IN CANADIAN FRENCH PHONOLOGY

Douglas C. Walker, Department of Linguistics, University of Ottawa, Ottawa, Canada K1N 6N5

Inspired by Delattre's question "Le mot est-il une entité phonétique en français?", many phoneticians and phonologists have investigated the status of the word in French phonology. While there is considerable caution expressed by the majority of researchers, the clear tendency is to minimize the role of the word, and to emphasize its subordinate status within the phonological phrase. Recently, however, evidence from phonotactic patterning and from morphophonological investigations, as well as from more specifically phonetic domains, has begun to re-establish the importance of word boundaries in French phonology.

In this paper, I will examine four types of allophonic variation in an informal variety of Canadian French (vowel laxing, lowering of /ɛ/, backing of /a/, and assibilation of apical stops). Each of these processes is sensitive either to the presence or the absence of word boundaries. This lower-level phonetic evidence, when coupled with more abstract data, allows us to re-affirm the importance of word boundaries, and the notion of "word", in French phonology.

## THE SOUND SHAPE OF LANGUAGE IN ALL ITS FACETS

Linda R. Waugh, Cornell University, Ithaca, New York, USA

It has long been recognized that language has 'double articulation': units with meaning are composed of units (distinctive features, phonemes, syllables) without meaning, whose only significance lies in their 'mere otherness'. However, the speech sound as a whole is an artifact endowed with many different functions, only one of which is distinctiveness. In particular, there are in addition redundant, configurative, expressive, and physiognomic features, each of which have a function of their own and none of which evidence 'double articulation'. In addition, the distinctive features evidence the tendency for immediate signification and autonomous significance, as shown by sound symbolism, by the role sounds play in magic (e.g. glossolalia), in language play (verbal games), in poetry (where the sounds become a focus of attention in their own right and where they are one of the constitutive devices of the sequence), and in 'word affinities' (identity of form between words, which affect the meaning and the history of the words - evidenced for example by 'phonesthemes'). It is concluded that phonology and phonetics are both currently being defined too narrowly - being confined to distinctiveness - and that the ever-occurring balance between mediacy ('double articulation') and immediacy (e.g. sound symbolism) for the distinctive features must be taken into account if we are to understand language structure and language change and if we are to be able to interpret our results in speech perception, child language acquisition, etc. correctly.

Reference

Jakobson, Roman & Linda R. Waugh (1979): The Sound Shape of Language, Bloomington, Ind.: Indiana University Press.

A PRELIMINARY STUDY OF DISTINCTIVE FEATURES AND THEIR  
CORRELATIONS IN STANDARD CHINESE

Zong-ji Wu, Institute of Linguistics, Chinese Academy of  
Social Sciences, Peking, China

This paper attempts to find the distinctive features of Standard Chinese according to the traditional classification in Chinese phonology, and three DF matrices are given for the SC vowels, consonants and tones. It also proposes an "N-binary" concept, according to the dialectic relations between binary and N-ary classifications, and provides a set of phonological correlation patterns revealing in this way the quantitative changes in SC phonemes.

Subjects

As the DF must be adequate for characterizing important phonetic differences between languages, we adopt several extra features from the traditional taxonomy of Chinese phonology in choosing the features for SC, i.e., open/closed and spread/protruded for vowels; aspirated/non-aspirated for consonants, and rising/falling, level/concave, high/low for tones.

Since a segment in the sequence of speech cannot be represented merely by two opposite features, and since the allophones between phonemes are varied almost continuously, we suggest here three patterns to designate these correlations. In the pattern of 7 vowels, 4 pairs of DF are distributed in a triangular diagram to designate such quantitative sound changes relevant to the variations of tongue positioning, jaw-opening and lip-rounding. In consonants, 4 pairs of DF are used to construct a matrix, in which 24 consonants of SC are positioned, to show the quantitative and qualitative changes. As for the tones of SC, 4 tones are sited at each corner of a quadrangle to build up a interwoven network of 16 combinations, in which the allotones of tone sandhi are shown.

## THE ROLE OF NEUTRALIZATION IN THE MECHANISM OF PHONOLOGICAL CHANGES

V.C. Zhuravlev, Institute of Linguistics, Moscow, USSR

The proposed conception solves some problems and explains general laws of diachronic phonology: "system pressure", "empty squares", catalysis, limitation of allophonic variation, etc.

Subjects

Having posed the corresponding formulae:  $a \rightarrow b+c$  (1);  $a \times b \rightarrow c$  (2), Polivanov reduced the empirical experience to two main types of sound changes - 1) divergence and 2) convergence. The problem of close interconnection between them was set: convergence as a rule is accompanied by divergence, and vice versa. Jakobson having brought the formulae together, proposed the combined formula of phonological sound changes:  $A_1:B_1 \rightarrow A_2:B_2$  (3). The case of Polivanov's 1 or 2 formula presupposes the appearance of a new opposition or disappearance of an old one. Another type of phonological changes was discovered. The opposition is preserved but the relationship between its members has been changed. The change of the phoneme turned out to be interconnected with the opposition. The necessity to solve these and other problems of diachronic phonology makes us look at the phenomenon of neutralization of phonological opposition at the present synchronic stage. Any neutralization may be expressed by the following combined formula:  $\frac{a:b}{Pr} \rightarrow \frac{c}{Pn}$  (4).

The power of the phonological opposition and the power of neutralization can be calculated:  $F^O = k \frac{d}{n}$ ,  $F^n = q \frac{n}{d}$ , where  $d$  is the number of differentiation positions (position of maximum differentiation),  $n$  the number of positions of neutralization (weak position). By means of coefficients (the number of correlated pairs -  $k$ , and the number of neutralized pairs -  $q$ ) the investigated opposition is included into the system of related oppositions - into the correlations. The comparison of the combined neutralized formula with Polivanov's convergence and divergence formulae reveals the difference only in the dependency of strong and weak positions.

Conclusion

Phoneme convergence and divergence should obligatorily pass through the neutralization stage by means of a correlation between the numbers according to the formulae (5) and (6). Neutralization observed at the present synchronic stage may potentially be regarded as the way either to divergence or to convergence, i.e. the arrows in formula (4) may be two-directional.

NASAL SOUNDS IN DOGRI

Ved Kumari Ghai, Department of Sanskrit, University of Jammu

This paper attempts to discuss the phonemic status and distribution of nasal sounds in Dogri - an Indo Aryan language of NW India. As nasal sounds of Dogri have developed from Sanskrit, a brief description of the treatment of nasals in Sanskrit has also been given.

Dogri has five nasal stops. In most of the modern Indo Aryan languages only bilabial and dental nasals appear as phonemes, while velar, palatal and cerebral nasals occur as homorganic nasals before corresponding non-nasal stops. Dogri, on the other hand, has all five nasal stops as phonemes.

Besides this category, Dogri has three more sub-categories of sounds exhibiting some aspect of a nasalization process:

1. Non-phonemic nasalization of vowels due to the presence of a nasal stop in the environment.
2. Phonemic nasalization of a vowel.
3. Homorganic nasal before a stop.

The first sub-category is generally predictable and is, therefore, not represented in writing. A peripheral vowel preceding a nasal stop in monosyllabic words is nasalized. In disyllabic words where both syllables are open, the nasal stop of the second syllable nasalizes both vowels. A nasal stop preceded by a centralized vowel and followed by a peripheral vowel does not cause nasalization of the following vowel. The last two sub-categories are represented by Anusvara, which is phonetically actualized in two different forms: as a phonemically nasalized vowel in word-final position and before vowels, and as homorganic nasal before a stop if the preceding vowel is centralized.

## LA COMPLEXITE DES LATERALES EN FRANCO-PROVENÇAL

Fernande Krier, Romanisches Seminar der Universität Kiel,  
Kiel (République Fédérale d'Allemagne)

Le matériau sur lequel se fondent les données du présent exposé est constitué par des enregistrements de langage spontané, effectués dans plusieurs villages du Val d'Anniviers (Valais, Suisse).

A la notation, nous avons relevé quatre consonnes latérales, l'apico-dentale sonore [l], la dorso-palatale sonore [ʎ], l'apicale vélarisée sonore [ɰ] et, fait rare qui, à notre connaissance, n'est attesté en Europe qu'en islandais, en féroé, dans certains dialectes celtiques ainsi qu'en sarde septentrional, la latérale apico-dentale sourde [ɭ].

L'essentiel de notre exposé consistera à démontrer qu'il s'agit là bien de quatre unités distinctives, ce qui fait que dans le parler francoprovençal en question, les latérales participent, à côté des occlusives et des fricatives, à la corrélation sourde/sonore.

Références

- Lüdtke, H. (1953): "Il sistema consonantico del sardo logudorese", Orbis 2, 411-422.
- Martinet, A. (1955): Economie des changements phonétiques, Berne: Francke.
- Martinet, A. (1956): La description phonologique avec application au parler franco-provençal d'Hauteville (Savoie), Genève: Droz, Paris: Minard.
- Ternes, E. (1973): The Phonemic Analysis of Scottish Gaelic, Hambourg: Buske.
- Werner, O. (1963): "Aspiration und stimmlose Nasale / Liquide im phonologischen System des Färingischen", Phonetica 9, 79-107.

THE VOICE-VOICELESS CONTRAST IN IRISH SONORANTS

Dónall P. Ó Baoill, Institiúid Teangeolaíochta Éireann, Dublin

The analysis of the sonorant consonants in Irish presents many intriguing problems for the linguist, for phonological theory and for the theory of universals of language. There are two developments which are worthy of consideration and I outline them briefly below.

- (i) The first problem has to do with the process of lenition which in general changes all stops to fricatives. The process is complicated by the fact that certain nasals and liquid consonants participate in the lenition process as well. In this case, one type of nasal/liquid is converted into another. This poses many problems for distinctive feature theory and the writing of rules.
- (ii) Secondly, the occurrence of voiceless nasals/liquids under certain conditions in Irish have to be dealt with. On the phonetic surface contrasts occur between the voiced and the voiceless type but the question to be answered is: do the same contrasts occur at a more abstract level? For example, the dialect of Irish to be discussed has eight phonemic voiced nasals (4 palatalized, 4 velarized). When the future morpheme or past participle ending is attached to words ending in nasals, the nasals tend to be devoiced. Thus, a contrast between the future tense and the present subjunctive is one of voiced-voiceless. If the contrast is phonemic, we have 16 nasal phonemes. These problems will be discussed, and suggestions made about what constitutes an appropriate analysis in this case.

/ʈs/: A VOICELESS UNASPIRATED EMPHATIC ALVEOLAR AFFRICATE

Edward Y. Odisho, Al-Mustansiriyah University, Baghdad, Iraq

As far as I can tell there is no mention in phonetic literature of a linguistic unit with the following description: a voiceless unaspirated emphatic alveolar affricate, to be transcribed, hereafter, as /ʈs/. This paper reports the existence of /ʈs/ in the Neo-Aramaic language spoken by the Assyrians in Iraq. It is pertinent to point out that the language has three other affricates namely /tʃ<sup>h</sup>/, /tʃ/ and /dʒ/ which represent aspirated, unaspirated and voiced palato-alveolar affricates, respectively.

It is worth mentioning that /ʈs/ has no plain counterpart in Neo-Aramaic, therefore one wonders how we have ascribed the features 'emphatic' and 'unaspirated' to the sound concerned. Spectrographic evidence shows that with /ʈs/, F<sub>1</sub> and F<sub>2</sub> behave exactly in the same manner as with other well-established emphatics both in Neo-Aramaic and Arabic, in that F<sub>1</sub> is raised while F<sub>2</sub> is lowered so as to achieve drastic approximation. In so far as the attribute 'unaspirated' is concerned, this is partly based on the auditory quality of /ʈs/ when compared with the German /ts/, and partly on a comparison with the Aramaic /tʃ/ and /tʃ<sup>h</sup>/<sup>1</sup> (for which the term 'aspiration' is broadly used to embrace both frication and aspiration occurring consecutively). The latter comparison shows that the aspiration phase of /ʈs/ is nearer in magnitude to that of /tʃ/ than to that of /tʃ<sup>h</sup>/. This phonetic similarity tempts one to envisage that /ʈs/ has possibly emerged in the system to function as the emphatic counterpart of /tʃ/, the shift in place of articulation being attributed to the availability of better chances for anchoring the tip/blade at the alveolar zone rather than at the palato-alveolar zone. Such anchoring is required to counter the tendency to tamper with the primary articulation under the pressure of the backing gesture, a manoeuvre that is necessary for the execution of the secondary articulation, i.e. pharyngealization.

---

1) For the phonetic details on /tʃ/ and /tʃ<sup>h</sup>/, see my paper in Journal of the International Phonetic Association 7, 1977.



QUICHEAN (MAYAN) GLOTTALIZED AND NON-GLOTTALIZED STOPS:  
 A PHONETIC STUDY WITH IMPLICATIONS FOR PHONOLOGICAL UNIVERSALS

Sandra Pinkerton, Phonology Laboratory, Department of Linguistics,  
 University of California, Berkeley, California

Investigators have noted that ejectives exhibit a preference for back articulations while implosives exhibit the opposite preference. A counterexample to Greenberg's implicational hierarchies for ejectives and implosives has been offered from the Quichean languages which have a glottalized set of stops consisting phonologically of "ʔb, t', k', ʔq". This counterexample and the lack of information about the phonetic nature of glottalized stops, particularly uvulars, led to the present phonetic study of Quichean glottalized and non-glottalized stops.

5 Quichean languages were investigated in Guatemala using portable equipment to get intra-oral air pressure and audio recordings. 27 male subjects were recorded, 15 of these from rural and urban K'ekchi speaking areas. The inventory consisted of 10 tokens each of 16 real language minimal pairs containing the stop contrasts for bilabial, alveolar, velar and uvular places of articulation in word initial and medial positions.

The extent of the phonetic variation across these languages shows that the identification of stops as "glottalized" by no means indicates their phonetic nature. Bilabial implosive variants are: ɓ, a voiced, negative pressure implosive; ɓ̥, a voiced, zero pressure implosive; b, a voiced, non-glottalized variant; p<, a voiceless implosive. Alveolar implosive variants are: d̥, a voiced, negative pressure implosive; t<, a voiceless implosive. The dialectal variation for the glottalized uvular stop in K'ekchi (Carcha K'ekchi - q' in all word positions; Chamelco K'ekchi - q< in all word positions; Coban K'ekchi - q' word initially, q< intervocalically) suggests that further work is needed to determine if there is a necessary phonetic relationship between the ejective and the implosive variants. The difference between voiceless and voiced implosives in these languages suggests that the best generalization about place of articulation preferences for glottalized stops is that voiced glottalized stops have a preference for front articulations and voiceless glottalized stops have a preference for back articulations.

## LINGUISTIC ATTRIBUTES OF RETROFLEX ʀ IN PIGNASCO

Gladys E. Saunders, University of Virginia, Department of French and General Linguistics, Charlottesville, Virginia 22903, USA

The nature of retroflex ʀ and its interaction with non-retroflex resonants is examined in Pignasco (a Gallo-Italian dialect spoken in western Liguria). It is shown that retroflex ʀ manifests a number of phonological characteristics that differentiate it from its non-retroflex congener (e.g. its 'resistance to contiguous palatal sounds; its restriction to weak syllable position; its correlation with the velar [ŋ] in certain morphological paradigms - cf. [bo:ʀa]-[boŋ] 'good', fem. sing. and masc. sing., respectively). It is argued that the retroflex ʀ in Pignasco cannot be considered as merely a variant of non-retroflex r (as a superficial analysis would lead one to maintain); rather, it must be defined in terms of the complexities of its phonetic properties along with universal notions of occurrence. (An adequate analysis of a dialect requires the dialectologist to take these factors into consideration.) Recent research on the behavior of retroflex consonants in other languages (cf. Bhat 1974; Stevens and Blumstein 1977) together with studies on the acquisition of ʀ in children (cf. Wode 1977) support the argument.

References

- Bhat, D.N.S. (1974): "Retroflexion and retraction", JPh 2, 233-237.  
 Stevens, K.N. and S.E. Blumstein (1975): "Quantal aspects of consonant production and perception: a study of retroflex stop consonants", JPh 3, 215-233.  
 Wode, H. (1977): "The L2 acquisition of /r/", Phonetica 34, 200-271.

UNIVERSAL AND LANGUAGE SPECIFIC TRAITS IN THE SYSTEM OF SOUND FEATURES

Albertas Steponavičius, University of Vilnius, USSR

This paper deals with hierarchies of distinctive features (DFs) and phonological oppositions. The set-up of hierarchies must be such that oppositions of a higher rank comprise oppositions of a lower rank. It follows from this that subclasses of different classes of phonemes are not structurally and functionally identical and must be set up independently, irrespective of the possible identity of the anthropophonic nature of their DFs. Classes of phonemes may also be separated into subclasses by more than one pair of DFs at a time. The main distinction to be made is that between consonants and vowels, with both liquids and glides classified as consonants. This primary distinction is expressed by means of two pairs of features, consonantal vs. nonconsonantal, and vocalic vs. nonvocalic, in view of the possible presence in some languages of items to be specified as /-con, -voc/. Of consonantal modal features primary importance should be attached to the features obstruent vs. nonobstruent, and sonant vs. nonsonant. The next pairs of modal features which must be classified among the primary and universal ones are stop vs. nonstop, and fricative vs. nonfricative. The consonantal distinction nasal vs. nonnasal, though language universal, is secondary from the point of view of particular languages in that it is usually relevant only in the subsystem of sonants. In determining the degree of the universality of the consonantal features according to place of articulation, a sharp distinction should be made between the so-called active and passive organs of speech. According to the participation of the active organs of speech labial, apical, and dorsal series of phonemes may be distinguished, specified by means of the following universal (or near-universal) DFs: apical vs. nonapical, labial vs. nonlabial, and dorsal vs. nondorsal. Further local specifications of consonants according to points in the stationary part of the vocal tract are highly language specific. Vocalic features of aperture are universal in that all the known languages have at least two vowel heights, the most regular type being three heights. The utilization of the primary vocalic features front vs. back, and rounded vs. unrounded is rather language specific, though separate typologies may be established.

## CLICKS, AND THEIR ROLE IN THE EVOLUTION OF LANGUAGE

Roman Stopa, Kraków, Poland

Gestures as well as their abbreviations in form of incomplete clicks (i.e. clicks alone without any back element, which in clickblocks is an essential part of the compound) cannot constitute any consistent and hierarchically organised system. Here the role of clickblocks as that of synthetizers and classifiers of the experienced situation appears to be essential. They can be considered as the direct predecessors of our labial, dental and lateral ranges of phonemes as they already have - though sometimes only roughly outlined - the linguistically so important features of a determined place and mode of formation: When a clickblock is transformed into an expiratory (or a clicklike, i.e. ejective, injective or disjunctive) phoneme, then the front part of the clickblock, e.g. /k' determines the place of articulation and the back part of it determines the mode - or the way - of the articulation. The clickblocks perform the 3 functions of language, the vowel with its suprasegmental features being the exponent of the expressive, the consonantal back element of a clickblock constituting the indicator of the communicative function, and the click itself, while integrating these elements and appreciating the whole of the situation as to its value for man's organism, symbolises the situation and opposes its sign to all the other clicking signs of experience.

This role of clicks in creating a certain system of symbols which reflect man's references to the interesting elements of his experiences leads to considering their place among all linguistic symbols, and especially, among all the sounds of human speech.

A survey of different click types in various languages will be given.

CHARACTERISTICS OF CONSONANT PRODUCTION/DEVELOPMENT  
FOR PRE-ADOLESCENT CHILDREN

W.S. Brown, Jr., IASCP and Department of Speech, University of Florida, Gainesville, Florida 32611

A considerable amount of supraglottal air pressure ( $P_{10}$ ) data has been generated to describe and quantify consonant production of adult speakers. On the contrary, only a meager amount of  $P_{10}$  data has been reported for children's speech. The present study collected  $P_{10}$  data for 120 normal school age children ranging in age from 5-10 years of age (10 males and 10 females were recorded for each age group).  $P_{10}$  was recorded via a polyethelene sensing tube placed through the corner of the mouth extending into the posterior portion of the oral cavity (behind the point of consonant articulatory constriction). The pressure sensing tube was in turn connected to a differential pressure transducer, the signal amplified and graphically displayed on one channel of an oscillographic recorder. Peak  $P_{10}$  values were obtained from the children repeating syllables (embedded in a carrier phrase) containing the stop-plosive pairs /p, b/ and /t, d/, and the continuant pair /s, z/ in a variety of syllabic positions. The  $P_{10}$  data from the children speakers were compared to an adult "model" of consonant production including: (1) overall air pressure values; (2) voice/voiceless consonant distinction; consonant class distinction; (3) effect of syllabic position; and (4) constancy of production. The results indicated that the pre-adolescent children have developed consonant productions that are nearly identical to mature adult speakers even as early as five years of age. These results will be discussed in terms of language acquisition and development in children.

## TYPES OF SYLLABLE DIVISION AMONG RUSSIAN CHILDREN OF DIFFERENT AGE GROUPS AND MODERN THEORIES REGARDING THE SYLLABLE

N.I. Lepskaya, E.N. Vinarskaya and G.M. Bogomazov, Moscow University, Moscow, USSR

The syllable structure of speech is formed gradually during the first ten years of a child's life under the influence of a fuller acquisition of the native language. Russian children of different age groups intuitively divide words into syllables in accordance with different language models.

The first step is to divide a word into syllabic segments, consisting of a sequence of sounds with increasing sonority. These segments are functionally indivisible (the model of the open syllable). Then the gradual increase in vocabulary and the mastering of the sound structure of words lead to the second step, i.e. these segments are turned into phonetic sequences, the most productive of which is the model of the closed syllable. A child's gradual understanding of the morphological structure of words becomes more fully reflected in his syllable division and leads to the third step, i.e. the morphological principle of syllable division.

All three types of syllable division are true, primarily of the initial, dynamically expressed part of a word.

Children of older age groups use all the above mentioned types of syllable division. This means that newly acquired skills do not completely oust earlier acquired skills. Taking these facts into consideration, it may be assumed that the above mentioned syllable models exist in the speech of adults.

The existence of a number of syllable models in language in accordance with different types of syllable division may account for the diverse theories in linguistics.

## PHONOLOGICAL STRUCTURE OF SPEECH ADDRESSED TO CHILDREN

Linda Shockey, Department of Languages and Linguistics, University of Essex, Wivenhoe Park, Colchester, England, and Z. S. Bond, School of Hearing and Speech Sciences, Ohio University, Athens, USA

An area of study which aims to clarify children's acquisition of their first language concerns the nature of speech addressed to children. The consensus seems to be that adults attempt to clarify and/or simplify the structure of their language when speaking to children. Although the exact function of adult simplification is still controversial, it may well serve to provide a language-learning child with a corpus of systematic and grammatically correct primary language data from which he can generalize the rules of his language with greater facility.

In the present study we investigated the phonological structure of child-directed utterances in comparison with the phonological structure of adult-directed utterances. In order to make an explicit comparison possible, we selected four phonological processes common in casual conversation and calculated their frequency of occurrence vs. their potential occurrence in adult-adult conversation and in adult-child conversation.

Mother-child and mother-adult conversations were recorded in an informal setting. Each child was given a puzzle which served as the topic of conversation; the adult-adult conversations dealt with informal topics. All the participants in the study were long-term residents of Colchester or the surrounding area.

The recorded conversation samples were analyzed for the ratios between the actual occurrence and the potential occurrence (i.e. when the structural description of a rule is met) of four phonological processes which tend to neutralize the distinction between lexical items. The phonological processes were:

- 1)  $t \rightarrow ? / \_ \#$  (loss of oral contact for /t/)
- 2)  $t \rightarrow \emptyset / \_ s \#$  (final cluster simplification)
- 3)  $t \# j \rightarrow t_j$  (affrication)
- 4)  $\delta \rightarrow \emptyset /$  apical continuant  $\# \_ \_$  ( $\delta$  loss)

Contrary to expectation, mothers were found to use a more reduced style of speech, characterized by a liberal use of common phonological processes, with their children than with adults.

These results raise the following question: How do children acquire the full representation of lexical items when they seldom hear it from their mothers?

## INHERENT STRUCTURE OF SEGMENTS: EVIDENCE FROM NATURAL EXPERIMENTS

M.E. Solberg, Department of Linguistics, M.I.T., Cambridge, USA

It has been assumed that a theory which includes only a set of distinctive features cannot adequately characterize the inherent structure of segments, since certain feature conjunctions are more likely than others. Several proposals have included a hierarchization of features as a partial solution to this problem. We describe a general experimental paradigm which permits us to bring ontogenetic evidence to bear on the issue of inherent structure. Application of the method to a specific case indicates that ontogeny may provide unequivocal evidence for feature ordering.

Quechua has three obstruent series: /p/, /t/, /k/, /ç/, /q/; /p<sup>h</sup>/, /t<sup>h</sup>/, /k<sup>h</sup>/, /ç<sup>h</sup>/, /q<sup>h</sup>/; /p<sup>ʔ</sup>/, /t<sup>ʔ</sup>/, /k<sup>ʔ</sup>/, /ç<sup>ʔ</sup>/, /q<sup>ʔ</sup>/. This symmetrical system constitutes a natural experiment in which we can isolate the features for aspiration and glottalization, while holding constant all other feature values. We recorded 250 hours of dialogue between 10 monolingual children (aged 1;4 to 5;1) and interlocutors in an Andean village. Three-hour samples were collected at monthly intervals for periods up to 22 months. For each subject we made five tests of the hypothesis that aspiration developed before glottalization by examining /p<sup>h</sup>/ vis-a-vis /p<sup>ʔ</sup>/, /t<sup>h</sup>/ and /t<sup>ʔ</sup>/, etc..

We found that for all pairs the development of C<sup>ʔ</sup> implied the development of C<sup>h</sup>, and that C<sup>h</sup> implied C. Additional experimental and naturalistic data collected from a larger sample of subjects five years after the initial study revealed no counterevidence. Moreover, when we look at less conservative dialects of Quechua we find that development predicts the change which has occurred in those dialects with respect to the laryngeal subsystem.

The Quechua result is especially interesting because the frequency of glottalized obstruents in texts and mature dialogue is significantly greater than the frequency of aspirated obstruents. When we examined the frequency of these ten obstruents in the speech which mature interlocutors addressed to the subjects, we found a frequency reversal at critical junctures in development.

Our results suggest the possibility that the feature ordering we found in Quechua may be universal to the species. The examination of evidence from additional natural experiments may be expected to corroborate or reveal additional inherent structure.



## MOTOR ANALYSIS OF INFANT SOUND

Jeannette M. van der Stelt and Florina J. Koopmans-van Beinum,  
Institute of Phonetic Sciences, University of Amsterdam,  
Amsterdam, Netherlands

In literature a phonological approach to the sounds of children in the first year of life is common, although the period is said to be prelinguistic (why not preadult?).

In this paper we report on a new way of infant sound analysis. We do not listen to linguistic elements in the sounds. We relate the sounds to events in the speech apparatus, noting separately respiration, phonation and articulation. A special system of symbols has been developed.

Knowledge of the infant's anatomy, physiology and development is indispensable.

Non-crying sounds of two male infants have been analysed from birth to eight months.

This way of analysing the infant sound production makes it possible to give a precise definition of babbling.

The philosophy behind this approach is the opinion that speech is in the first place a sequence of movements. A behaviouristic study of speech will be quite revealing.

In speech development the child learns to relate (speech) movements to meaning. Parent-child interaction is essential for this learning process.

References

- Van der Stelt, J.M. and F.J. Koopmans-van Beinum (1978): "Note on motor analysis of infant sound", Proceedings V, Institute of Phonetic Sciences, University of Amsterdam (to appear).
- Koopmans-van Beinum, F.J. and J.M. Van der Stelt (1978): "Early Stages in Infant Speech Development", Proceedings V, Institute of Phonetic Sciences, University of Amsterdam (to appear).

## ON SOME BASIC PRINCIPLES IN CHILD PHONOLOGY

Jaroslava Pačesová, Department of Phonetics, J. E. Purkyně University, Brno, Czechoslovakia

In her paper the author attempts to present the most outstanding operating principles which seem to govern the learning process at the phonological level. In agreement with Jakobson, her theory of phonemic development makes essentially three claims:

- the sound system of a child has structure in the same way that adult phonology has structure; though simplified at the early stages of language development, it has similar entities, similar patterns of variation and distribution and, in addition, shows regular patterns of substitution for adult phonemes;
- the mastering of a phonemic repertory can best be described in terms of the successive acquisition of increasingly differentiated oppositions of distinctive features;
- a universal pattern of development exists which is also mirrored in the distribution of feature contrasts among languages generally.

The early acquisition of minimal vocalism and minimal consonantism reveals the basic principle, i.e. the principle of maximum contrast, viz. close versus open, low versus high, front versus back, oral versus nasal and accounts for the stability and wide distribution of the vowels /a/, /i/, /u/ and of the consonants /p/, /m/, /t/, /n/.

Next there are the following principles, which appear to operate in child language and in languages generally:

the priority of unmarkedness over markedness, occlusivity over fricativity, labiality and/or alveolarity over velarity and simplicity over complexity. Their manifestation is shown in the precedence (with regard to both stability and distribution)

- of unmarked phonemes as opposed to marked ones;
- of stop phonemes as opposed to fricative ones;
- of front consonants as opposed to those whose place of articulation is the velum;
- of simple fricatives as opposed to laterals and vibrants;
- of simple vowels as opposed to vowel chains, whether diphthongal or hiatic.

## LANGUAGE ACQUISITION AND PHONETIC SIMILARITY

Henning Wode, Englisches Seminar der Universität Kiel,  
 W.-Deutschland

In the past, linguistic and phonetic theories have been thought helpful to interpret language acquisition data, in particular L1 acquisition. The more sophisticated the model, the more sophisticated the interpretational possibilities offered to students in language acquisition. This hope seems unwarranted because currently available theories have been developed for fully fledged adult languages and not for learners, children and/or adults. The inadequacy of this approach derives from the fact that learners very often react to properties of the target language which do not figure in the linguist's formal description of the particular language at all, or which do so much less prominently than they deserve in view of their importance for language acquisition. It is suggested that language acquisition be explored as to what insights it may offer for linguistic theorizing. As for phonetics/phonology, it is suggested that acquisition data, in particular from L2 acquisition, will throw light on the problem of phonetic similarity, and that, perhaps, transcription systems should be revised to accord with such insights. Consider the L2 acquisition of the various types of "r". In general, L2 learners replace the L2 targets by the closest equivalent of their L1 repertoire. If not interfered with by teaching or orthography, such learners will first replace the uvular [ʀ] or [ʁ] by [χ] or [h], even if their L1 "r" is the retroflex [ɽ], the frictionless [ʃ] or the rolled alveolar [r]; and the retroflex L2 [ɽ] or the frictionless [ʃ] will be replaced by [w], even if the learner's L1 "r" is [ʁ ʀ r]. Obviously, to the learner [ʁ ʀ] are more similar to [χ h] than to [ɽ ʃ ʃ], and [ɽ ʃ] more similar to [w] than to [ʁ ʀ r]. Since such observations are not anecdotal but systematic in the sense that they are specific to all learners of a specific acquisitional type, these phonetic regularities should be reflected in the phonetic transcription. As for the various "r"'s this can easily be done via appropriate feature specifications. This approach will be extended to other phonetic elements.

## SIMULATION DYNAMIQUE EN TEMPS REEL DES PHENOMENES DE PRODUCTION DE LA PAROLE

A. Bourjault, Laboratoire d'Automatique, ENSCMB, Besançon, France

Dans le but de développer un outil de recherches destiné à mieux connaître les mécanismes de production de la parole naturelle, nous avons réalisé et testé un système hybride pour simuler en temps réel ces phénomènes.

En considérant le conduit vocal comme un système mécanique globalement générateur de signaux, continûment déformable dans l'espace et le temps par le locuteur, nous avons établi un modèle mathématique effectivement dynamique. Les équations rendent compte de l'évolution de la pression et de la vitesse de l'air au sein de l'appareil phonatoire, sous l'action des variations des aires transversales. Ainsi, les sources d'excitation (source glottique, sources de bruit ...) ne sont plus traitées à part; elles apparaissent dès lors que sont simulées les conditions qui permettent leur existence: ouverture et fermeture de la glotte, constriction, occlusion.

Les équations sont résolues sur le Simulateur Analogique Modulaire rapide S.A.M. (construit au Laboratoire), les variables de commande (les fonctions d'aire) étant fournies au modèle par un calculateur numérique.

Une première série d'expériences a permis de réaliser les 12 voyelles orales du français et des groupements V.C.V. (/apa/, /ada/, /ara/, /ala/, /aza/...). La forme particulièrement simple de la programmation permet d'étudier divers paramètres comme la durée, l'allure des transitions, les points d'articulation, etc... Il est possible d'envisager également d'autres applications, comme l'analyse directe par synthèse ou la synthèse des langues à tons.

#### Bibliographie

- André, P., A. Bourjault, A. Chevillard et J.M. Henrioud (1975): "Calculateur analogique rapide pour la simulation en temps réel des phénomènes de phonation", Symposium International "Simulation'75", Zurich.
- Bourjault, A. et A. Chevillard (1976): "Le problème des sources dans la simulation dynamique du tractus vocal", 7èmes Journées d'étude sur la parole, Nancy.
- Bourjault, A., A. Chevillard et F. Lhote (1976): "Hybrid simulation of the vocal tract", 8ème Congrès Association Internationale pour le Calcul Analogique et Hybride, "Simulation of systems", Delft.

## TEXT-TO-SPEECH CONVERSION BY RULE AND A PRACTICAL APPLICATION

Peter B. Deneš, Mark Y. Liberman and Joseph P. Olive, Bell Laboratories, Murray Hill, New Jersey 07974, USA

A system for the rule synthesis of voice answerback sentences for telephone directory-assistance purposes is described. The sentences have the form "The number for (Joe Snooks) of (518 Oaklands Avenue) is (345-6789)". Research on such a system offers the attractions of a genuine practical application for rule synthesis. It combines a non-trivial text-to-speech conversion task for the large numbers of names and addresses involved, yet avoids many of the unknowns associated with the synthesis of general English text because only a single carrier sentence is used. Also, evaluation of comprehensibility can be more realistic, using genuine users with a communication task, rather than laboratory subjects.

The task is performed in two steps. First the text is converted by rule into phonetic transcription, including stress and segment durations. The spelling-decoding involves a limited morphological analysis and a set of context-sensitive rewriting rules. Stress is assigned by a simplified version of the principles in Liberman and Prince (1977). A small dictionary of orthographically exceptional words is maintained. Durations are assigned by a set of rules which take into account the segment, its segmental context, the local stress pattern, and constituent structure.

The output of the above process serves as input to the stage of the synthesis process in which intonation is determined and the acoustic wave is calculated. The pitch contour is obtained by selecting and adjusting one of several stored contours. The acoustic wave is calculated by dyadic concatenation of vocal tract area function segments: the concatenation is based on a matrix of phoneme transitions stored as vocal tract area parameters.

The computer program implementing the first of the above two steps runs on a PDP11/45 several times faster than the associated speech time. In tests of randomly selected telephone directory entries, 91% of all entries were given a "correct" phonetic transcription and stress pattern. The second step also runs in real time, using a specially wired vocal tract area function synthesiser.

Reference

Liberman, M.Y. and A. Prince (1977): "On stress and linguistic rhythm", Linguistic Inquiry 8.2, 249-336.

## PRELUDE A LA PAROLE-ORDINATEUR: LE DICTIONNAIRE EVAGRAPHIQUE

Etienne Emerit, Université de Lille III, France, et Institut de Linguistique et de Phonétique de l'Université d'Alger, Algérie

Après huit années de recherche expérimentale sur synthétiseur à formants "EVA III", et la mise au point d'une méthode universelle de recherche des logatomes optimaux, applicable à toutes les langues, l'auteur présente son "Dictionnaire évagraphique".

Ce dictionnaire se compose d'épreuves photographiques étalonnées en temps, fréquences et intensités, et tensions électriques correspondantes, pour piloter les modules du synthétiseur. Ces épreuves, obtenues sur l'"Evascope" inventé par l'auteur, permettent d'optimiser le dessin des évagrammes sans passer par la préanalyse, et d'autre part rendent possible la mise en mémoire-ordinateur des signes de parole.

En effet, au prix de l'addition de quelques paramètres supplémentaires relativement simples (sans modification des paramètres existants) tels que simulation des attaques de bruit coloré, de l'interchangeabilité de la source vocale, et de la suppression de la diaphonie des sources de bruit, les 95% d'intelligibilité actuelle pourraient être portés à 100% par composition automatique et lecture par ordinateur, infiniment plus précise et fiable que la lecture originelle du dessin des évagrammes par système électromécanique.

## PHONETIC MODELING - THEORY AND APPLICATION

Georg Heike, Institute of Phonetics, University of Cologne

Some of the most important goals of phonetic research should be (a) to explain the universal principles of speech processes, (b) to describe the language specific solutions thereof, (c) to build machines that help man to communicate with other machines or via machines with other people. This work can only partly be done by describing phenomena. The main tool seems to be the use of explicitly defined models which can be implemented on computers in order to test hypotheses.

A phonetic model of speech communication should be differentiated within two planes: (1) along a horizontal axis from speaking to hearing, and (2) by different vertical levels of phonetic and linguistic information. Concerning the first plane the model should consist of the following main parts which are interrelated with each other: (1) a parametric acoustico-genetic synthesizer, (2) an acoustic analyzer with parameter extraction, (3) a phonetic processor for the control of synthesis, analysis and recognition. Some of the main problems to be solved are: coarticulation, assimilation and compensation in synthesis, extraction of articulatory parameters in acoustic analysis, segmentation and classification in recognition. While the acoustico-genetic synthesizer simulated on a computer already works, we hope to present the results of combining synthesis with analysis. The control of an articulatory model by acoustic parameters and simulation of compensation should be possible.

## SYNTHESIS OF ESTONIAN LANGUAGE

Eugen Kynnap, Institute of Cybernetics, Academy of Sciences of the Estonian SSR, Tallinn, ESSR

The work described here presents results of synthesizing Estonian by means of a terminal synthesizer with serial and parallel connection of filters. The synthesizer has two branches: one with a buzz generator to synthesize vowels, and the other with a hiss generator to synthesize unvoiced consonants. To synthesize voiced consonants both branches are used at the same time. A low-frequency generator of complex-form tension, elaborated for this purpose, acts as the buzz source. It is possible to generate pulses of any form. A special digital control system was created to control the analog circuits of the synthesizer. There are 12 controllable parameters, which were not all constantly used. This system allows to observe and alter the tracks of all control parameters during the experiments. The transitions of parameters are chosen linear.

In Estonian 32 phonemes, including 9 vowels, are distinguished. Voiceless plosives /p,t,k/ have three cues: 1) the burst of noise, 2) the silence, and 3) the transition of the formant of vowels preceding or following them. The semi-voiced counterparts /b,d,g/ have the same cues, only they have a tone impulse in the initial phase of their production, their noise burst is weaker and longer and silence shorter. The duration of all plosives in the initial position of syllables is shorter than in the final position. The fricatives /s,h,f/ were synthesized only by means of the noise components. The lower cutoff frequency of noise, when forming /s/, depends on the phoneme, which stands before it. The consonants /l,n,s,t/ and conventionally /d/ have both palatalized and unpalatalized forms, not distinguished in a written text. The palatalization is performed by means of /i/-like transitions. When the palatalized consonant occurs in the final position of a syllable, the i-like transition is attached to the formants of the unpalatalized counterpart of the phoneme, producing a syllable with a very weak initial /i/.



## OPTIMAL INTONATION CONTOURS FOR POLISH SPEECH SYNTHESIS

Wojciech Majewski, Wojciech Myślecki and Janusz Zalewski,

Institute of Telecommunication and Acoustics, Technical University of Wrocław, Wrocław, Poland

This paper is focused on the selection of interrogative and declarative synthetic intonation contours which in the opinion of listeners provide the most naturally sounding statements and yes-no questions. In contrast to the previous studies (1, 2) that were utilized as a basis for the present investigation, the synthetic stimuli varied not only in the fundamental frequency but were generated by means of a set of rules which permitted a simultaneous control of pitch, intensity and duration.

#### Procedure

Synthetic stimuli were generated by rule on a computer simulated formant series synthesizer. The experimental material consisted of two phrases: CVCV ("jola") and VCVCV ("uleje"), on which different intonation contours were superposed. The fundamental frequency ( $F_0$ ) pattern was obtained from the glottal excitation amplitude ( $A_0$ ) pattern by means of the following rule:

$$F_0 = FO \frac{A_0 + a}{1 + a}$$

where FO is the  $F_0$  target value (Hz), and a is a numerical coefficient. A phrase intonation contour,  $F_{0c}$ , was obtained by multiplying  $F_0$  by the intonation function  $F_k$ , approximated by a linear function. The stimuli were tape-recorded, randomized and presented to a group of listeners who evaluated the stimuli for naturalness.

#### Conclusions

The results of the experiments permitted to establish the simple rules generating the intonation contours for interrogative and declarative short phrases of Polish synthetic speech. An important conclusion resulting from the experiments is that the realization of interrogative and declarative intonation takes place in a relatively short final segment of a phrase and because of that it is not necessary to calculate the intonation function for the total duration of a phrase.

#### References

1. Majewski, W. and R. Blasdel (1969): "Influence of fundamental frequency...", JASA 45, 450-457.
2. Studdert-Kennedy, M. and K. Hadding (1973): "Auditory and linguistic processes...", L&S 16, 293-313.

## EXAMPLES OF SOME SYNTHESIZED HUNGARIAN SENTENCES

Antti Sovijärvi, Department of Phonetics, University of Helsinki, Vironkatu 1 B, 00170 Helsinki 17, Finland

As the basis for my speech synthesis investigation of Hungarian I have made use of the spectrographic material which accrued during the course of my analysis of the intervocalic alveolo-palatals /c, ʃ, ɲ/ (Sovijärvi, 1975a, 1976). In my first sentence synthesis tests I used unmodified parameter values for the subordinate phases of the realisations of the sounds in accordance with this data. I concluded that although it was unnecessary to use five subordinate phases for any phoneme occurring - this holds true for both geminates and single consonants as well as for long and short vowels - taking only two phases into consideration would have been insufficient. After numerous experiments I arrived at the compromise that for model synthesis of each sound realisation utilization of four subordinate phases is a methodologically practicable strategy at this stage.

The program used to control the OVE IIIb synthesizer from the HP2000 computer was written by Marita and Göte Nyman (1978), according to specifications based on our discussions.

In conjunction with my presentation I shall offer for scrutiny examples of variant renditions of six of the sentences synthesized. On the basis of the matrix for one example sentence Atyámat gyászolom [ 'ʰa:ta:maʃ 'q̣ja:solom], some important matters of principle will be brought up which were encountered during the calculation of the parametric values of the sound phases.

Mr. Reijo Aulanko has assisted me in these experiments both as operator and as research assistant.

#### References

- Sovijärvi, A.A.I. (1975a): "Results of the Spectrographic Analysis of Alveolo-Palatals in Hungarian", paper read at the Eighth International Congress of Phonetic Sciences (Leeds 1975).
- Sovijärvi, A.A.I. (1976): "Unkarin kielen liudentuneiden dentaalien spektrografisen analyysin tuloksia", in Fonetiikan paperit - Helsinki 1975, Publications of the Institute of Phonetics - University of Helsinki, 27: 89-114.
- Nyman, Marita and Göte Nyman (1978): "OVE IIIb - syntetisaattorin ohjausohjelman käyttöopas" (= Guide book for using the feeding program for the OVE IIIb synthesizer), Mimeographed Series of the Department of Phonetics - University of Helsinki, Teaching Material 1.

## SPEECH SYNTHESIS FROM INTERPOLATED LOG-AREA CODED TRANSITIONS

Hans Werner Strube, Drittes Physikalisches Institut, Universität Göttingen, Bürgerstrasse 42-44, D-3400 Göttingen, F.R. Germany

A speech synthesis system is described, based on concatenation of elements from a pool of 48 stationary sound segments together with many (up to 1322) sound transitions. The speech signals are coded by 13 log-area-ratios,  $\log(\text{pitch})$ ,  $\log(\text{power})$  and two binary parameters switching the noise and pulse generators. Output is generated by a computer-controlled hardware synthesizer (Strube 1977). The stationary sounds are coded by a single parameter-frame and the transitions by two frames only, taken from real speech. Intermediate frames are restored by interpolation during synthesis. Direct linear interpolation of the above parameters is in most cases a fairly good choice compared to other possibilities. The transition boundaries to be stored were determined using the spectra, the parameter curves and a subjectively optimized fit of straight-line trains to the curves.

The transition table is addressed through a (37 x 37)-matrix with first and second phoneme as row and column index, also containing the transition length. Thus the same frame pairs may be used for different transitions, also in opposite time sequence. When a transition is not yet in the table, single phonemes are concatenated; for many pairs, this is even good enough. The quasi-stationary part of a sound is either also interpolated or inserted discontinuously as a constant portion. Treatment of different phonemes, excitation, timing, and intrinsic pitch are controlled by a sound-property table.

Synthesis occurs in real time during input-text evaluation. Input is in ASCII characters, closely matching the IPA transcription. Pitch is controlled by numbers in the input text, whereas duration and intensity are given by the program. Pitch changes are smoothed and intrinsic pitch is added by the program. Investigations in automatic intonation rules (J. Kretschmar 1978) and intrinsic pitch in German are in progress. Results and examples will be presented.

#### References

- Kretschmar, J. (1978): "Untersuchungen zum Tonhöhenverlauf deutscher Sätze für die Sprachsynthese", Fortschritte der Akustik-DAGA '78, Berlin: VDE-Verlag, 455-458.
- Strube, H.W. (1977): "Synthesis part of a 'log area ratio' vocoder in analog hardware", IEEE Trans. ASSP-25, 387-391.

UN MODULE DE TRAITEMENT DU TEXTE ECRIT EN FRANCAIS EN VUE  
DE LA SYNTHÈSE AUTOMATIQUE PAR DIPHONÈME

Daniel Teil et Bernard Prouts, LIMSI - CNRS 91406 Orsay, France

Ce module, destiné à remplacer celui implanté dans notre système de synthèse (Teil et al., 1972), détermine à partir d'un texte écrit en français orthographié les éléments nécessaires à sa synthèse par diphonèmes: suite des phonèmes, valeurs du pitch, valeurs temporelles.

Le programme orthoépique se résume à l'exploitation d'un lexique contenant indifféremment les règles de prononciation et les exceptions. Dans ce lexique sont inclus des marques de liaison et des marques prosodiques. Le traitement des homographes qui ne peut être fait que par analyse syntaxique et sémantique n'est pas traité actuellement.

L'interdépendance de la prosodie avec la syntaxe n'étant pas encore clairement établie nous avons choisi une méthode lexicale de découpage du texte en groupes prosodiques. Les ponctuations déterminent les coupes fortes; les coupes faibles séparent des groupements de mots établis en fonction d'une liste de mots "outils" (Choppy et al., 1975).

La courbe mélodique est calculée à partir d'un schéma en dent de scie appliqué au niveau du groupe mélodique et au niveau de la phrase en fonction de la ponctuation. Le rythme est actuellement traité de façon assez sommaire.

Conclusion

Les résultats obtenus sont assez encourageants et nous incitent à parfaire l'algorithme surtout au niveau de la prosodie.

Références

- C. Choppy, J.S. Liénard et D. Teil (1975): "Un algorithme de prosodie automatique sans analyse syntaxique", 6èmes Journées d'Etude sur la Parole, Toulouse.
- D. Teil, J. Sapaly et J.S. Liénard (1972): "Conception et réalisation d'un terminal vocal d'ordinateur à vocabulaire illimité et réponse immédiate", 3rd International Congress on Data Processing in Europe, Salzbourg.

SELECTION OF GLOTTAL EXCITATION PARAMETERS OPTIMIZING  
THE NATURALNESS OF SYNTHETIC SPEECH

Janusz Zalewski and Wojciech Myślecki, Institute of Telecommunication and Acoustics, Technical University of Wrocław, Wrocław, Poland

The shape and periodicity of source excitation influences the naturalness of synthetic speech (1,2). In the present study these problems were investigated for short phrases of Polish synthetic speech.

Procedure, Experiments and Results

Glottal pulses were shaped by means of time functions previously examined by Rosenberg (2). Amplitude and frequency of the glottal excitation were controlled by a set of simple rules. All synthetic utterances were generated by a digital series formant synthesizer and were subjectively evaluated by means of an A-B test. To obtain an interval preference scale, Thurstone's model V of comparative judgment was accepted and Mosteller's least squares solution (3) was used.

The goal of the experiments was to determine optimal pulse shape function, optimal relative opening  $t_o$  and closing  $t_c$  times, and optimal amplitude  $A_D$  and frequency  $F_D$  of fine pitch deviation. In the first experiment optimal  $t_o, t_c$  combinations for five examined pulse shape functions  $f_A, \dots, f_E^2$  were established. In the next experiment these functions were compared for naturalness and it was found that the best was the trigonometric function  $f_C$  with  $t_o = 0.41$ ,  $t_c = 0.2$ . The optimal deviation parameters were  $A_D = 1.7$  Hz and  $F_D = 6$  Hz.

Conclusions

The results of the experiments have shown that pulse shape and  $t_o, t_c$  values strongly influence the naturalness of synthetic phrases. The obtained preference scores for investigated phrases, synthesized applying various pulse shape functions, are quite similar to Rosenberg's results (2). It was also found that, for each pulse shape, besides the optimal  $t_o, t_c$  pair, there was a distinctively different second pair of  $t_o, t_c$  which provided comparable naturalness.

References

1. Zalewski, J. and W. Myślecki (1975): "Research on the selection of  $F_o$  for the optimal synthesis of vowels", 8 Int. Cong. Phon. Sc., Leeds.
2. Rosenberg, E.A. (1971): "Effect of glottal pulse shape on the quality of speech from parallel formant synthesizer", JASA 49.
3. Mosteller, F. (1951): "Remarks on the method of paired comparisons - the least squares solution", Psychometrica 16.

THE PROSODY OF GRAMMAR AND THE GRAMMAR OF PROSODY<sup>1</sup>

Olga Akhmanova and L'udmila Minaeva, Moscow State University,  
Moscow, USSR

The 'Prosody of Grammar'

"Grammar" when used to mean "Syntax" is primarily prosody. The fallacy of "paper syntax", the utter untenability of vociferous but scientifically unfounded pronouncements concerning the "grammaticality" or "ungrammaticality" of ridiculous strings of written slovoforms is demonstrated by referring the reader to an unusually convincing piece of linguistic material (Young et al., 1970, 306) and one of the more recent reports of a relevant investigation (Lehiste, 1977).

The 'Grammar of Prosody'

"Grammar" can also be used to mean "general facts" (as against "special" ones, Henry Sweet). In this sense the "Grammar" of prosody pervades the whole of language. The less generally known aspect of this enormous field is "lexicological phonetics" - a new branch of phonetics whose aim is to prove the objective existence of words as units of language, analyze and explain the expression plane of lexical categories.

References

- R.E. Young, A.L. Becker, and K.L. Pike (1970): Rhetoric: Discovery and Change, New York: Harcourt, Brace and World, Inc.
- Lehiste, I. (1977): "Isochrony reconsidered", JPh 5, 253-263.

---

(1) We have taken the liberty to adopt the "paradox-pattern" which was so successfully introduced by professor Roman Jakobson.

## INTERRELATION OF RHYTHM AND OTHER COMPONENTS OF INTONATION

A.M. Antipova, Institute of Foreign Languages, Moscow

Intonation is understood here as a close unity of speech melody, voice quality, sentence stress, temporal characteristics and rhythm. Hence the object of investigation is the character of changes in all the components of intonation under the influence of change in rhythm.

A piece of prose and a poem served as the experimental material. Five speakers read the material in two ways: first as they personally felt it should be read, then with increased rhythmicity (i.e. with greater emphasis on rhythm).

The results are as follows:

1. Emphasized rhythmicity increases the tendency towards monotone at the perceptual level which is determined by narrow intervals in the fluctuation of the fundamental frequency. This tendency is more pronounced in prose. In verse, however, the slowing down of tempo adds greatly to the impression of monotony.
2. Stresses are intensified which is determined by the increase in intensity and time.
3. The tempo is slowed down due to the increased duration of phonation and pauses.
4. The voice quality is changed. In the present material it becomes clearer and softer.

Consequently, the quantitative change (increase) in rhythm leads to the changes in other components. As a result, a qualitatively new pattern is produced which expresses a different meaning.

PHENOMENES DE RUPTURE ET DE NON RUPTURE EN FRANCAIS PARLE  
EXPLORATION DE CERTAINES RELATIONS ENTRE STRUCTURES PROSODIQUES  
ET SYNTACTICO-SEMANTIQUES. UNE CONTRIBUTION A LA PHONOSYNTAXE.

Anne Bergheaud, Laboratoire de Phonétique, Département de Recherches Linguistiques, Université Paris VII, Paris, France.

A l'origine, notre sujet d'étude portait sur la liaison en français parlé en tant qu'indice d'une "proximité" des éléments au niveau profond de l'analyse syntaxique.

Sans renier notre intérêt pour ce sujet, les obstacles que présente toute étude isolée de la liaison nous ont conduite à l'envisager dans un cadre différent et à formuler l'hypothèse suivante : certains points des séquences parlées en français présentent :

- soit un ensemble de faits phoniques de "RUPTURE" (pause, au moins "perçue", coup de glotte, rupture de rythme, écart intonatif, modification d'intensité, absence de liaison) présents en totalité ou en partie.
- soit un ensemble de faits phoniques de "NON RUPTURE" (ni pause, ni coup de glotte, rythme régulier, intonation et intensité non altérées, liaison éventuelle).

L'analyse instrumentale ainsi que les tests psychoacoustiques que nous menons actuellement, ont produit de premiers résultats assez significatifs, et nous pensons être en mesure d'en donner de plus complets dans quelque temps.

Cette nouvelle approche présente, selon nous, trois avantages :

- La liaison ou son absence font partie désormais d'un faisceau de faits phoniques qui se placent dans le domaine général de la structuration de la prosodie.
- Une telle étude entre dans un cadre théorique "phonosyntaxique" qui traite de l'agencement des phénomènes prosodiques en relation avec des structures syntaxiques.
- Il est alors possible de soulever des problèmes de syntaxe "fine" qui abordent des questions de structure sémantique relevant de l'"attitudinal meaning", questions qui peuvent ainsi trouver un traitement cohérent et un cadre théorique.



## ZUM PROBLEM DER SEGMENTIERUNG DER FREQUENZKONTUREN

L.P. Blochina, das Moskauer Institut für Fremdsprachen

Die Frage über die Methoden der Frequenzkonturenanalyse für die Differenzierung der phonologischen und phonetischen Information bleibt bis jetzt ungelöst und wird noch diskutiert. Die meisten Linguisten sind der Meinung, dass die Tonstufenanalyse, die die Segmentierung der Frequenzkontur in diskrete Einheiten implizit voraussetzt, vorzuziehen ist. Es gibt auch verschiedene Meinungen betreffs der Methoden der Segmentierung der Frequenzkontur, und die am meisten verbreitete von ihnen ist die Betrachtung der Frequenzkonturen als solche, die aus diskreten Teilen besteht. Dieser Meinung nach, gehören die diskreten Teile zu verschiedenen Tonstufen, deren Zahl von 2 bis 5 variiert. Die Methoden der Aussonderung der Tonstufen die dabei verwendet werden, sind entweder rein intuitiv; oder für ihre Aussonderung wird die Information von der Inhaltsstufe herangezogen, oder die vertikale Segmentierung wird rein formell verwirklicht.

Im vorliegenden Vortrag wird der Algorithmus der Bildung der Frequenzkontur auf Grund der statistischen Angaben dargelegt und die Ergebnisse seiner Approbation an Hand des Materiales der russischen, englischen und deutschen Aussage- und Fragesätze gezeigt. Auf Grund der durchgeführten vertikalen Segmentierung werden zwei Methoden der horizontalen Segmentierung der Frequenzkonturen vorgeschlagen.

Die ausgesonderten Segmente werden als diskrete Teile der Frequenzkonturen betrachtet, die Frequenzintervalle (Tonbrüche) werden an ihren Grenzen als akustische Merkmale dieser Frequenzkonturen interpretiert. Die mit Hilfe des vorgeschlagenen Algorithmus ausgesonderten akustischen Merkmale wurden im Prozesse der auditiven Analyse der Frequenzkonturen der Aussage- und Fragesätze von den Informanten nachgeprüft und bestätigt.

Zwecks der Schaffung günstiger Bedingungen für den Vergleich der Frequenzkonturen verschiedener Sprachen wird der primäre Algorithmus der Bildung der Frequenzkonturen durch den Algorithmus der Berechnung des Abstandes zwischen den Tonstufen ergänzt.

THE SIGNIFICANCE OF NON-PHONEMATIC COMPONENTS OF THE SOUND CHAIN  
FOR THE SUPRASEGMENTAL LEVEL OF THE LANGUAGE SYSTEM (BASED ON  
RUSSIAN LANGUAGE MATERIAL)

G.M. Bogomazov and R.F. Paufoshima, Institute of Russian Language,  
Moscow, USSR

In the Russian language, as in many others, there are some elements which are not significant for the phonological system. The elements of the speech chain which are dealt with are: inserted vowels, glottal stops and so on. Though these elements have no significance in themselves, they are significant for the other language levels, for example, for the suprasegmental level.

Thus, the inserted vowels participate in the formation of syllables in speech, contributing in this way to the creation of the specific features of the syllable structure of a language. This is expressed in the tendency to realize the consonant groups with the help of inserted vowels which may be observed in the Russian literary language and in a number of Russian dialects. This is connected with the domination in these language systems of open syllables. Another type of realization of consonant groups (without inserted vowels) demonstrates the frequency in speech of closed syllables and this tendency may be observed in the Ukrainian and in some North Russian dialects.

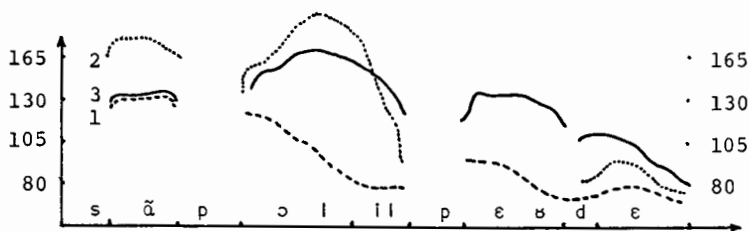
A non-phonematic glottal stop ([ʔ]) in the Russian language and in a number of its dialects marks the initial stressed vowel. In this way, it underlines the beginning of a word.

In order to assess the significance of such speech elements, it is necessary to accept the point of view that our perception follows the different levels of the sound chain.

"PERDU OU PAS?" UNE ETUDE SUR LA CONTRIBUTION DE L'INTONATION  
A LA STRUCTURATION DE L'ENONCE EN FRANCAIS ET DE SES RAPPORTS  
AVEC L'ORDRE DES MOTS

Georges Boulakia, Dpt. de Recherches Linguistiques,  
Université Paris 7

Dans le cadre d'études plus générales entreprises par notre groupe de recherche de l'Université Paris 7, sur les rapports des structures intonatives et syntaxico-sémantiques il est discuté dans cette communication du rapport de l'intonation (réduite à la mélodie) et de l'ordre des mots dans des phrases françaises telles que "sans Paul ils perdaient" ou "10 mètres plus loin le train déraillait". Dans ces phrases le complément peut être en tête, mais dans tous les cas il y a 2 structures syntaxiques possibles distinguées par 2 patrons intonatifs (en particulier localisation d'un sommet). Ces phrases ne sont pas syntaxiquement ambiguës, mais dans leur réalisation la distinction intonative peut être neutralisée, ce qui au cours de tests de reconnaissance de phrases naturelles ou synthétiques isolées, provoquera une confusion.



Phrases 'naturelles' "Sans Paul ils perdaient"

- 1: pas perdu = (reconnu)
- 2: perdu (réponse 'partielle') = reconnu
- 3: perdu ('récit') = peu reconnu

MALE AND FEMALE INTONATION: A CAUSE OF BRITISH-AMERICAN  
MISUNDERSTANDING

Ruth M. Brend, Michigan State University, East Lansing, Mi., USA

This study will attempt to partly answer the question of the basis of cross-cultural misunderstandings between speakers of "one" language - English. That is, it will attempt to explain, for example, why Americans sound brusque or angry to Britishers, when they do not sound so to other Americans, and why British men often sound effeminate to Americans. A continuation of the author's earlier study on male and female intonation in American English, this new study includes male and female patterns in both general American and (a variety of) British English. These patterns will be compared and contrasted, using primarily the O'Connor and Arnold framework.

References

- Brend, Ruth M. (1972): "Male-Female Intonation Patterns in American English", Proc. Phon. 7, The Hague, Mouton, 866-870.
- O'Connor, J.D. and G.F. Arnold (1973): Intonation of Colloquial English, London: Longman Group Ltd.

CONTRIBUTION A L'ETUDE DE LA PROSODIE GENERATIVE: STRUCTURES  
TEMPORELLES DES PHRASES ENONCIATIVES SIMPLES ET ETENDUES EN  
FRANCAIS

G. Caelen, Laboratoire C.E.R.F.I.A., Université Paul Sabatier,  
Toulouse, France

Cette étude s'intègre dans une recherche plus vaste consacrée aux structures prosodiques de la phrase énonciative en français, menée à partir d'une analyse acoustique non perceptuelle des trois paramètres: fréquence fondamentale, énergie, durée, dont les évolutions générales au sein des énoncés sont formalisées en un système de réécriture. Cet article en particulier propose une systématisation nouvelle des évolutions temporelles dans une perspective générativiste.

Ce modèle génère des énoncés temporels selon deux types d'unités structurales déterminant deux formes de réécriture possibles, à l'aide de deux règles.

Nous établissons par ailleurs une distinction entre une syntaxe textuelle et une syntaxe prosodique, la dernière n'étant pas la réalisation concrète de la première. Toutes deux, en leur domaine respectif et selon leur système spécifique de règles, proposent une mise en relation des unités linguistiques dont les frontières dans les deux systèmes peuvent ou non coïncider.

## ON AN ALGORITHMIC STUDY OF ENGLISH INTONATION

L.A. Canter, M.A. Sokolova and A.P. Tchizhov, Moscow State Pedagogical Institute, English Department (USSR)

This paper represents the first attempt to apply an algorithmic method to the study of English intonation. The method involves a new computer-assisted technique of acoustic analysis. It advantageously replaces the heuristic method, hitherto in extensive use. The algorithmic method makes it feasible to take into account the correlation of initial parameters and to give a quantitative estimate of their significance for differentiation of opposed intonation types.

The purpose of this investigation is a computerized search for one optimal acoustic distinctive feature with reference to a general linguistic dichotomy - statement/question. The experiment was designed to analyze fundamental frequency ( $F_0$ ) in the utterance "You knew [...] ≠ You knew [?]" . Each test phrase was pronounced in an appropriate context by 13 subjects, all speakers of British English. 26 pitch contours were obtained (13 statements and 13 questions, respectively). 25 of these, correctly identified by listeners, were used for further intonographic analysis. 8 initial parameters of the experimental material were analyzed: maximal and minimal  $F_0$  values within each syllable,  $F_0$  at the starting point,  $F_0$  at the end point,  $F_0$  at the last turning point, maximal  $F_0$  value between the starting point and the last turning point.

The acoustic distinctive feature conception as a linear combination (weighted sum) of all the initial parameters makes it possible to regard it as vector  $\theta = \langle P_1, \dots, P_T \rangle$ , where  $T$  represents the number of initial parameters,  $P_1, \dots, P_T$  are weight coefficients. If vector  $\theta$  differentiates the opposed pitch contours it can be considered as a distinctive feature, while each  $|P_i|$  ( $1 \leq i \leq T$ ) value can be viewed as estimates of the initial parameters' significance. Vector  $\theta$  was computer determined in a manner whereby all the pitch curves' values of statements in reference to  $\theta$  were positive and question pitch values were negative.

#### Conclusion

For the first time a linear combination of the initial parameters, ensuring optimal statement/question differentiation in English, was found.

## PRESENTATION D'UNE METHODE DE STYLISATION PROSODIQUE

Albert di Cristo, Robert Espesser et Yukihito Nishinuma,  
Institut de Phonétique, Université de Provence, Aix-en-Provence

Le but de cette communication est de présenter une méthode de stylisation des tracés de Fo fondée sur des critères acoustiques et perceptuels, en vue d'une application à des recherches sur l'intonation du français qui sont orientées, plus particulièrement, vers des études intonosyntaxiques.

Il est bien connu que l'analyse des structures intonatives ne peut procéder d'une interprétation directe des variations de Fo. Nous savons que ces variations reflètent diverses contraintes et qu'elles ne sont pas perçues en l'état par l'auditeur. Il importe donc de procéder à une stylisation des tracés acoustiques, en vue de dégager les variations prosodiques qui reflètent la compétence linguistique du locuteur.

Les méthodes de stylisation prosodique qui ont été élaborées jusqu'à présent sont peu nombreuses et souvent très incomplètes. Il convient, cependant, de citer les travaux très intéressants réalisés par les chercheurs de l'Institut d'Eindhoven ('t Hart, Collier et Cohen) ainsi que les récentes tentatives de Thorsen (Copenhague) et de Takefuta (Université de Chiba).

La méthode que nous présentons consiste à procéder à l'effacement des variations microprosodiques et à la conversion des données acoustiques en données perceptuelles. La conversion perceptuelle tient compte des seuils psychoacoustiques de Fo, de durée et d'intensité, ainsi que de la perception interactive de ces paramètres.

La stylisation des tracés de Fo est opérée en plusieurs étapes, à l'aide d'un ordinateur. Elle comprend des procédures semi-automatiques (opérateurs de linéarisation, de translation, de parabolisation, etc...) et des procédures automatiques (application des seuils différentiels de durée de Fo, d'intensité, du seuil de glissando, normalisation, etc...).

La comparaison des tracés stylisés d'après cette méthode avec des tracés de parole réitérée (méthode de répétition: pa, pa, pa) permet de constater que la technique de stylisation employée se prête particulièrement bien à l'objet de notre recherche.

## FALLS AND RISES: MEANINGS AND UNIVERSALS

Alan Cruttenden, Department of General Linguistics, University of Manchester

A basic distinction between some type of fall and some type of rise exists in a majority of the world's languages. This distinction has been seen at different times as relevant to grammar, lexis, discourse or attitude. At a higher level of abstraction all such meanings of intonation have something in common: the meanings typically associated with falling tunes, e.g. 'finality' 'closed-listing' 'response-denying' 'dogmatic' appear to have a common factor which may be called STRONG; while those associated with rising tunes e.g. 'continuity' 'open-listing' 'response-requesting' 'deferential' appear to have a common factor which may be called WEAK.

In some languages the distinction between fall and rise is either not used at all or is used only peripherally. Languages of this kind have a compensating increase in their use of distinctions of pitch height. In such cases meanings conveyed in other languages by a distinction between rise and fall are conveyed by the height of the terminal pitch, which may involve a distinction between a fall to mid pitch and a fall to bottom pitch or a rise to mid pitch and a rise to high pitch.

The use of intonation in languages may thus be stated in terms of a number of intonation universals: (i) if a fall v. rise distinction is used for certain dimensions of meaning, the correlations of form and meaning will be predictable; (ii) use of the fall v. rise distinction to convey one dimension of meaning will imply its use for certain other specifiable dimensions; (iii) where the fall v. rise distinction is used for several dimensions of meaning, certain dimensions will predictably always overrule certain other dimensions; (iv) if the fall v. rise distinction is not used in a language, then the language will use a distinction of pitch height to convey dimensions of meaning associated with fall v. rise in other languages.



## RHYTHM IN MODERN GREEK

Rebecca M. Dauer, University of Edinburgh, Edinburgh, Scotland

Is Modern Greek a stress-timed or a syllable-timed language? This question is investigated through a comparison of Greek and English based mainly on readings of prose texts. The two languages are compared with respect to syllable lengths, rate of speaking and interstress intervals (or 'feet'). The two most important findings are that the average interstress interval is about the same in both Greek and English - about .5 second, and that the ratio of foot lengths in Greek increases in the proportion of 1 : 1.5 : 2 : 2.5 : 3 : 3.5 from one to six syllable feet.

This shows that foot lengths correlate with the number of syllables they contain, increasing by the addition of one unstressed syllable which has one half the quantitative value of a stressed syllable. It is mainly the alternation of vowel lengths within the foot that is important in Greek and establishes this rhythmic pattern. Thus, although we can say Modern Greek has a rhythm of alternation, it is neither a pure syllable-timed nor a pure stress-timed language.

## STYLISTIC LOAD OF PROSODIC FEATURES IN ENGLISH

Yuri A. Dubovsky, Minsk State Institute of Foreign Languages,  
Minsk, USSR

The object of this paper is to consider the effects brought about by modifications of the prosodic text structures and, ultimately, to present some evidence on the stylistic load of prosodic features in English.

Procedure of the experiment

Obtained by way of instrumental analysis, the prosodic features for four verbally identical but stylistically different English text types were synthesized - conversational informal (CI) and formal (CF) monologues, a public speech at a relatively big indoor meeting (PI) and a public speech at a very big gathering of people in the open air (PO). Each prosodic text structure was transformed so that it contained a) either tone, tempo or intensity features of one of the remaining three texts, b) three parameters from three different texts, thus forming complexes like a) tone CI + tempo CI + intensity CF, or b) tone CI + tempo CF + intensity PI. Forty listeners were instructed to state whether a text, recorded in a random sequence, was acceptable for English and if so, to give it some stylistic label.

Conclusion

The prosodic text structures have different degrees of tolerance to modifications to preserve their stylistic individuality. The behaviour of a prosodic parameter in the text is stylistically determined, with various correlations between text types and prosodic features.

The distinctive stylistic semantics of the text is created, on the prosodic level, by at least two prosodic parameters, the greatest functional load among which is carried by tone features.

IMPORTANCE RELATIVE DES PARAMETRES DE L'ACCENT (DUREE ET FREQUENCE FONDAMENTALE) DANS LA PERCEPTION DE L'EMPHASE

D. Duez, Institut de Phonétique, Domaine Universitaire, Grenoble et R. Carré, Laboratoire de la Communication Parlée, Equipe de Recherche Associée au C.N.R.S., E.N.S.E.R.G. 38031 Grenoble Cédex

L'expérience décrite porte sur un extrait d'un discours politique de Pompidou (1973) offrant des variations élevées de durée et de fréquence fondamentale dans la réalisation de l'accent, ces variations correspondant à une recherche d'expressivité. Des modifications opérées sur ces deux paramètres en utilisant un système d'analyse-synthèse à codage prédictif, ont permis de mettre en évidence leurs rôles respectifs dans la perception de l'emphase.

Expérimentation

La voyelle originale accentuée a une durée de 35 cs (soit 4 fois la durée de la voyelle inaccentuée correspondante), l'écart tonal entre cette voyelle et la voyelle précédente est de 5 demi-tons. A partir de la phrase originale, 14 phrases ont été obtenues par réduction de la durée de la voyelle et/ou réduction de l'écart tonal. Les phrases synthétiques ont été présentées à l'écoute de 10 auditeurs chargés de noter le caractère emphatique ou non emphatique des échantillons.

Résultats

La fréquence fondamentale ne joue pratiquement aucun rôle dans la perception de l'emphase. Cette dernière est fonction de la durée de la voyelle, le seuil étant situé à 20 cs environ. Au dessous, il semble que l'on puisse retrouver une certaine emphase, à condition de porter l'écart tonal à 8 demi-tons. Mais alors, la phrase obtenue semble moins "naturelle".

Conclusions

Notre première expérience pour déterminer les paramètres de l'emphase donne des résultats encourageants et met en évidence le rôle essentiel de la durée.

Nous allons maintenant poursuivre notre recherche à partir de phrases différentes, de voyelles différentes, faire varier l'intensité...

Références

- Duez, D. (1978): Essais sur la prosodie du discours politique, Thèse 3ème Cycle, Paris.
- Katwijk, A. (1969): "The perception of stress", Institute for Perception Research, Eindhoven, Annual Progress Report 4, 69-73.

## TONEME PATTERN CONTOURS IN NORWEGIAN

K. Fintoft, Department of Linguistics, University of Trondheim, Norway

The purpose of the investigation is to study how the toneme patterns (the fundamental frequency in minimal tonemic pairs) change from one part of the country to the other. The investigation has been restricted to disyllabic words of the type /--V:CV/. Recordings have been made of about 1000 adult subjects from about 450 different places. For each speaker about 20 tracings of each toneme have been analyzed. For each speaker/place the positions of the maxima and minima on the toneme curves have been calculated relative to the duration of the sequence /V:C/. On the basis of the average curves for the different places, the realization of the two tonemes have been characterized by

1. the relative position of the main  $F_0$  peak or the peak in the stressed syllable.
2. The degree of similarity between the two toneme curves, given as a correlation coefficient.
3. the constants  $\alpha$ ,  $\omega$  and  $\phi$  of a damped sine function  $y = e^{\alpha x} \sin (\omega t + \phi)$ .

Maps are prepared indicating different values of these characterizations. In this way the dynamic aspect of the toneme curves is easily studied. From the different contour maps the old communication routes between Eastern and Western Norway are clearly seen. In some areas we see that the two toneme patterns correspond to toneme patterns in two different regions. In some areas we see how the patterns gradually change in such a way that the toneme curves coincide. When we examine the toneme curves in this way, it seems clear that the realization of the tonemes reflects the relationship between different geographical areas and the main communication routes in former days.

## COMPARISON OF WORD ACCENT FEATURES IN ENGLISH AND JAPANESE

Hiroya Fujisaki, Keikichi Hirose, University of Tokyo, Tokyo, and Miyoko Sugitō, Osaka Shō-in Women's College, Osaka, Japan

The word accent in various languages displays both universal and language-specific characteristics. While it is known that the voice fundamental frequency is the primary feature both in English and in Japanese, the duration and vowel color are also known to be important in English. This paper presents a comparison of these features between disyllabic words of English ("permit", "record", "object", etc.) and those of Japanese ("ame") of the *Osaka* dialect.

Fundamental frequency contours ( $F_0$ -contours)

It has been shown that the characteristics of  $F_0$ -contours of Japanese words can be well represented by the onset and offset of the accent command, extracted from the  $F_0$ -contour on the basis of a functional model proposed by Fujisaki and Sudō (1971). The same model was applied here to the analysis of English words and proved to be equally valid. While a marked similarity can be observed between  $F_0$ -contour characteristics of English and Japanese in cases of both first-syllable accented and second-syllable accented, individual differences are much greater in the onset of the accent command for English words with an accented first syllable.

Segmental and syllabic durations

Segmental durations were measured on the speech waveform, and were used to analyze the effect of accent position on the syllabic duration. It was found that accentual changes in duration occur mainly in the second syllable in Japanese, while in English they tend to be complementary in the first and second syllables.

Formant frequencies of syllabic nuclei

Formant frequencies of syllabic nuclei were extracted from the frequency spectrum by the method of Analysis-by-Synthesis developed by Fujisaki et al. (1970). It was found that accentual changes in formant frequencies are much greater in some English words (e.g. "record") than in others (e.g. "permit"), while they are invariably quite small in Japanese words.

References

- Fujisaki, H. and H. Sudō (1971): "Synthesis by rule of prosodic features of connected Japanese," Proc. Acoust. 7 3, 133-136.
- Fujisaki, H. et al. (1970): "Analysis, normalization, and recognition of sustained Japanese vowels," JASJ 26, 152-154.

## A PERCEPTION TEST OF PROSODIC FEATURES IN STANDARD SERBO-CROATIAN

Jadranka Gvozdanović, Slavisch Seminarium, Universiteit van Amsterdam,  
Amsterdam, Holland

The paper discusses the possibility that in some cases more than one hypothesis can be formed concerning a grammatical description, and that none of these hypotheses are rejected.

A perception test of prosodic features in Standard Serbo-Croatian is described. Serbo-Croatian is a South Slavic language, in which the basic unit of prosody is the so-called prosodic word. Prosodic word boundaries are indicated by means of a non-rising pitch which is followed by a high pitch. Within a prosodic word, only rising pitch is followed by a high pitch. The end of each prosodic word is characterised by a non-rising pitch. One non-final syllable nucleus in a prosodic word has a distinctive rising vs. falling pitch, which coincides with the place of the accent. Acoustically, the [+rising] pitch equals a rising fundamental frequency which is followed by a high fundamental frequency in the next following syllable, whereas the [-rising] pitch equals a non-rising non-falling fundamental frequency which is followed by a low fundamental frequency. The syllables following the one with the distinctive pitch are acoustically falling, and those preceding it, non-rising non-falling. (This is valid for prosodic words spoken in isolation. Under the influence of sentence intonation, regular modifications occur.)

I did a perception test with native speakers of Standard Serbo-Croatian in order to establish phonetic correlates of the place of the accent in a prosodic word. The parameters of fundamental frequency contour (expressed as a percentage of the duration of the syllable nucleus prior to the occurrence of the peak), maximal value of the fundamental frequency (expressed in Herz, transformed in a logarithmic measure), and duration of the syllable nuclei (expressed in milliseconds), were correlated with perception data. There are two hypotheses which are not rejected by the test: 1) the syllable with the rising fundamental frequency, or in its absence the first syllable in a prosodic word, which has a non-rising non-falling fundamental frequency, is accented, or 2) the last syllable in a prosodic word which is characterised by a non-falling fundamental frequency is accented.

The possibility that the second hypothesis cannot be rejected can be seen as a source of language innovation. In the Standard Serbo-Croatian prosodic system, the [-rising] pitch could originally be accented only in the initial syllable of a prosodic word, whereas the [+rising] pitch could be accented in any non-final syllable. In new compounds, however, a [-rising] pitch can be perceived as accented even when occurring in a non-initial syllable.

## A POSSIBLE 'NON-AUTONOMOUS' PHONOLOGICAL UNIT IN NORWEGIAN

Lars Hellan, Linguistics Department, University of Trondheim, Norway

There is some evidence that one domain for tone-assignment rules in Norwegian is close to, but not identical to, the word as defined by syntactic or morphological criteria: this unit is a morphological/syntactic word combined with unstressed neighboring elements and may hence be called a phonological word. One example is units like br<sup>2</sup>enner-opp ('2' indicating tone 2) ('burns up'), where opp may be seen as contracted to brenner, which has tone 1 in isolation, inducing tone 2. A rule accounting for this fact can be naturally obtained as an expansion from a general tone-rule schema whose other expansions can apply to syntactic/morphological words. Another example is given in the contracted for-lit<sup>1</sup>en ('too small'). In isolation, lit<sup>2</sup>en has tone 2, but assuming that for lit<sup>1</sup>en here acts as a word with regard to the tone-rule, the 'change' is accounted for by the general rule that only word-initial syllables can have tone 2.

Given that this phonological word is created (formally, presumably, by a 'restructuring' process applying to some syntactic level of representation) specifically for the demands of phonological rules, it might conceivably be a highly 'autonomous' phonological unit, internally structured only with regard to phonemes, syllables, quantity and stress at the point where tone rules apply. As shown in Haugen 1967, however, tone rules require a very articulate morphological analysis in their input. One instructive example is that although bisyllabic words often have tone 2, there is a regular rule to the effect that when the second syllable is a morph representing the definite article (which is suffigated), be it in the form -en, -a or -et, then the word has tone 1. The only exceptions to this rule are even more indicative of the abstractness of the input to tone rules: they are words like g<sup>2</sup>ata, h<sup>2</sup>ytta, whose indefinite forms are g<sup>2</sup>ate ('street') and h<sup>2</sup>ytte ('cabin'), both bisyllabic, as opposed to the monosyllabic indefinite forms in the cases where we get tone 1. A simple segmentation of g<sup>2</sup>ata cannot bring this fact out.

Further demonstration of the lack of phonological autonomy of the 'phonological word' will be given, also drawing on stress and quantity assignment.

Reference: Haugen, E. (1967): "On the rules of Norwegian tonality", Lg. 43, 185-202.

## ON THE NATURE OF FALL RISE INTONATION IN ENGLISH

Anthony Hind, Dpt. de Recherches Linguistiques, Université Paris 7

In this paper I will be discussing two facts concerning fall-rise intonation in English which I have discovered while looking for facts for or against an interpretative theory of intonation.

I will show that an intonation pattern which I will call a "marked" form of fall-rise pattern could be treated in relation to "un-marked" fall-rise intonation as a point further along a continual scale of "fall-riseyness". Data from a different source, however, tends to show that this marked fall-rise is the result of an entirely different process and can be best explained as a method of caricature of the basic pattern. Fall-rise intonation may therefore itself be part of a discrete system of intonation and therefore be explainable within an interpretative theory of intonation but this process of caricature is obviously a non-discrete system of coding information and cannot therefore be treated in an interpretative theory of intonation.

A second fact is put forward which cannot be accounted for by an interpretative theory of the syntactic type presented in Hirst (1974). I show that final rise is not the only factor in the intonation pattern which gives the reading not just any to a sentence such as I won't speak to any doctor; altering that final rise into a fall does not necessarily change the interpretation of this sentence to not any at all as predicted by the syntactic theory.

#### Reference

Hirst, D.J. (1974): La levée de l'ambiguïté syntaxique par les traits intonatifs, thèse, Aix-en-Provence (non-publiée).



## PITCH FEATURES FOR TONE AND INTONATION

D.J. Hirst, C.N.R.S., Institut de Phonétique, Université de Provence, France

This paper discusses some aspects of what Chomsky and Halle (1968,IX) referred to as "the still quite open question of the systematic role of pitch contours or levels".

It is argued firstly that a derived phonetic representation will need to include features of pitch whether or not these features are part of the lexical specification. A given pitch pattern should thus be expected to have the same surface representation whatever the language in which it occurs.

Secondly it is argued that features such as (HIGH) and (LOW) are best considered as defining a pitch interval, i.e. as meaning "higher" or "lower" than the last value. This, however, necessitates at least one feature which is absolute for a given speaker and which we refer to as (RESET). It is shown that providing (LOW) defines a greater pitch interval than (HIGH), no further conventions are required to account for downstep and downdrift in terraced-level languages and that the same three features could account for up to four levels in discrete-level languages. Data from recent acoustic studies of Bambara (Diarra, 1976) and Japanese (Nishinuma, 1977) suggest that it is possible to further generalise and specify (HIGH) and (LOW) as a fixed percent increase/decrease for a given speaker.

Finally it is argued that a further pitch-feature is required (TONIC) to account for the fact that in some cases pitch intervals seem to be specified between two non-adjacent syllables and that the pitch of intervening syllables is subsequently interpolated.

#### References

- Chomsky, N. and M. Halle (1968): The Sound Pattern of English, New York: Harper & Row.
- Diarra, B. (1976): Etude acoustique et fonctionnelle des tons du bambara (Mali), (Doctoral dissertation: Université de Provence).
- Nishinuma, Y. (1977): Contribution à l'étude prosodique du japonais: accent et intonation. (Doctoral dissertation: Université de Provence).

THE ESTIMATION OF INTRINSIC  $F_0$ : A COMPARATIVE STUDY

D.J. Hirst, A. Di Cristo and Y. Nishinuma, Institut de Phonétique, Université de Provence, France

A large number of studies have been devoted to the question of the intrinsic frequency ( $F_{0i}$ ) of vowels in various different languages. These studies consistently indicate a strong inverse correlation between  $F_{0i}$  and the first formant of the vowel. The coefficient of determination ( $R^2$ ) between  $F_1$  and  $F_{0i}$  for the data given by Peterson and Barney (1952) is 0.85. Calculating the regression line from  $F_1$  to  $F_{0i}$  consequently gives a reasonably close estimation of  $F_{0i}$ . This estimation can be considerably improved if we take into account the second formant ( $F_2$ ), since we obtain an  $R^2$  of 0.922. An even better correlation is found between  $F_{0i}$  on the one hand and  $F_1$ ,  $F_2$  and  $\overline{F_0}$  (the mean  $F_0$  for each subject) on the other hand, ( $R^2 = 0.976$ ) for the data from 11 different authors on 6 different languages. The estimation from the multiple linear regression on these data is very close to the original data ( $r = 0.988$ ) and, although the correlation varies from author to author, in most cases the difference between the estimation and observed values rarely exceeds 2%.

A linear function  $F_{0i} = a_0 + a_1 \overline{F_0} + a_2 F_1 + a_3 F_2$  where  $a_0 = 20.166$ ,  $a_1 = 0.975$ ,  $a_2 = -0.034$ ,  $a_3 = -0.002$  provides a very reliable estimation of the intrinsic frequency of vowels which can consequently be used both in prosodic analysis and in automatic speech synthesis and recognition.

Reference

Peterson, G.E. and H.L. Barney (1952): "Control methods used in a study of vowels", JASA 24, 175-184.

## TEMPO EFFECTS ON THE DURATION OF JAPANESE VOWELS AND CONSONANTS

Yayoi Homma, Dept. of English, Osaka Gakuin Univ., Osaka, Japan

This paper attempts to observe how speech tempos influence the durational relationship between consonants and vowels in Japanese.

Subjects

Three speakers read a list of bisyllabic words in carrier phrases with three different speech tempos: slow, natural and fast. We measured the duration of the phrases, the test words, the voice onset time (VOT) of the initial stops, the first and second vowels, and the medial stop closure duration or frication duration. At the slow tempo, the lengthening ratio for each item was almost the same as that for the whole phrase. At the fast tempo, however, the timing structure was altered: accented first vowels were less reduced, but at the same time VOT was more reduced. In general, when comparing different combinations, it was found that when medial consonant duration was less reduced, the second vowel was more reduced, or reduced to the same extent as in other combinations, and when medial consonant duration was more reduced, or reduced to the same extent as other consonants, the second vowel was less reduced. Thus a fixed reduction ratio was kept for word duration and other parts of the phrase.

Conclusion

The present study supported the results of my previous paper (Homma, 1978a) which showed that in Japanese, given a certain number of moras, closure duration, VOT and vowel duration work together to obtain fixed word duration. VOT in Japanese has smaller values both for voicing lag and voicing lead than in other languages (Lisker and Abramson, 1964; Homma, 1978b). At the fast tempo, VOT was reduced more than other parts of the utterance. This may imply that VOT is not so important a cue to separate Japanese voiced and voiceless stops as in English (Port, 1977).

References

- Homma, Y. (1978a): "An acoustic study of Japanese stops: closure duration, Voice Onset Time and their relationship with vowel duration", unpublished.
- Homma, Y. (1978b): "Voice Onset Time in Japanese stops", unpublished.
- Lisker, L. and A.S. Abramson (1964): "A cross-language study of voicing in initial stops: acoustic measurements", Word 20, 384-422.
- Port, R.F. (1977): Influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words, Bloomington: Indiana University Linguistics Club.

## DIE RHYTHMISCHE GRUNDSTRUKTUR DES RUSSISCHEN WORTES

E. Jasová, Pädagogische Hochschule, Banská Bystrica

Diese Abhandlung gibt eine quantitative Charakteristik der grundlegenden rhythmischen Struktur des Wortes in der russischen Schriftsprache auf Grund der Distribution der betonten und unbetonten Silben in Wörtern und Formen, die einer Untersuchung russischer schriftlicher Texten im künstlerischen und wissenschaftlichen Sprachstil entstammen.

Das Thema

Der Gegenstand unserer Forschung sind zwei Grundebenen der Struktur des russischen Wortes und zwar die Silbenebene im Verhältnis zur Ebene der Akzentstelle. Auf der Grundlage einer Analyse versuchen wir, die Frequenz bestimmter rhythmischer Worttypen prozentuell auszuwerten und graphisch zu demonstrieren. Bei den 2-silbigen Wörtern, die im Russischen am häufigsten sind, finden wir in der Frequenz der zwei möglichen rhythmischen Typen: 2/1 und 2/2 (d.h. Wörter mit dem Akzent auf der ersten und der zweiten Silbe) ein gewisses Gleichgewicht. Bei den 3-silbigen Wörtern tritt der rhythmische Typ 3/2 (d.h. Wörter mit dem Akzent auf der zweiten Silbe) deutlich in den Vordergrund. Bei den 4-silbigen Wörtern sind zwei Typen: 4/2 und 4/3 (Wörter mit dem Akzent auf der zweiten und der dritten Silbe) die häufigsten. Bei den 5-silbigen Wörtern ist der folgende Typ am verbreitetsten: 5/3 (d.h. Wörter mit dem Akzent auf der dritten Silbe). Eine bestimmte Tendenz der prosodischen Belastung der mittleren Silbe des russischen Wortes erkennen wir bei den 6- und mehrsilbigen Wörtern, obwohl die Frequenz dieser Wörter im Russischen sehr niedrig ist.

Konklusion

Im allgemeinen können wir feststellen, dass sich in der Distribution des russischen Wortakzentes eine deutliche Tendenz zur Akzentuierung der mittleren Wortsilben abzeichnet.

Literaturhinweise

- Zlatoustova, L.V. (1975): "Rhythmic Structure Types in Russian Speech", in Auditory Analysis and Perception of Speech, G. Fant and M.A.A. Tatham (ed.), 477-487, London, New York, San Francisco: Academic Press.
- Bondarko, L.V., L.R. Zinder et A.S. Stern (1977): "Nekotorie statistitscheskie charakteristiki ruskoj retschi", in Sluch i retsch v norme i patologii, Leningrad, 3-16.

## STRESS AS THE BASIS OF THE SWEDISH ACCENT DISTINCTION

John T. Jensen, University of Ottawa, Ottawa, Ontario,  
Canada K1N 6N5

The Swedish accents have traditionally and in recent generative analyses been treated as prosodic features covering two or more syllables within a word. In standard Swedish, accent II has two phonetic pitch peaks corresponding to the primary and secondary stress, while accent I has a single peak corresponding to the primary stress. I propose a synchronic analysis in which stress is assigned to words by rules similar to those developed by Chomsky and Halle (1968) for English, with the phonological cycle accounting for stress subordination. A PITCH rule assigns pitch peaks to the primary and a following secondary stress, if any.

The main body of the paper focusses on four stress rules for Swedish. The cyclic COMPOUND rule assigns the stress pattern ...1...2... to compounds (and certain other constructions). The result undergoes the PITCH rule to receive a pitch contour perceived as accent II. The MAIN stress and RETRACTION rules account for the ...1...2... pattern of words like spé<sup>1</sup>gél<sup>2</sup> 'mirror', which also have accent II. The THEME stress rule operates in words like gorilla<sup>1</sup><sup>2</sup> 'gorilla', giving these a ...1...2... stress pattern and thus accent II. These rules must be extrinsically ordered as (1) THEME stress, (2) MAIN stress, (3) RETRACTION.

These rules are sufficient to explain the difference in the location of secondary stress (and hence in the shape of the pitch contour) of word pairs like jördandé<sup>1</sup><sup>2</sup> 'burying' and jördände<sup>1</sup><sup>2</sup> 'earth spirit'. The COMPOUND rule, properly formalized, predicts accent I for compound verbs of the type betála<sup>1</sup> 'pay', although the basis verb tála<sup>1</sup><sup>2</sup> 'speak' has accent II. This analysis of betála can be generalized to adjective phrases of the type för många<sup>1</sup> 'too many', which have accent I, although the adjective många 'many' in isolation has accent II.

The Swedish accents are best understood as a phonetic reflex of stress patterns and can be described by universal pitch features without language specific features like Linell's [+Accent II].

#### References

- Chomsky, N. and M. Halle (1968): "The Sound Pattern of English",  
New York: Harper and Row.
- Linell, P. (1972): "Remarks on Swedish Morphology", Reports from  
Uppsala University, Department of Linguistics 1.

## ETUDE DES FONCTIONS DISTINCTIVES DE LA PROSODIE DE L'ENONCIATION

Véra C. Kachkina, Université d'état de Voronej, URSS

La recherche expérimentale des caractéristiques prosodiques des sous-types de phrases énonciatives telles que la narration, le titre, la réponse, la déclaration, l'annonce et la nouvelle porte sur l'étude de leur fonction communicative. Le corpus expérimental a permis de dégager dans l'analyse auditive, en plus des sous-types à intonation "neutre", deux sous-types expressifs de la réponse implicative et de la déclaration emphatique.

Dans une nette optique de différenciation des termes de "fonction" et de "procédé" linguistique (Trubetzkoy, 1960, 254), les procédés délimitatifs nous ont permis de dégager les unités pertinentes, tandis que les culminateurs nous ont servi à relever les unités les plus importantes sémantiquement et à les opposer aux unités secondaires de la segmentation de la chaîne parlée.

L'analyse phonologique corrélative perceptuelle et acoustique (Jakobson et al., 1952) des culminateurs des indices correspondants nous a permis de dégager dans sept oppositions phonologiques binaires variables la régularité de la répartition des sous-types des phrases énonciatives dans le sens des fonctions communicatives: explicative, appellative, expressive.

#### Conclusion

Le sens communicatif des sous-types de phrases énonciatives selon leurs caractéristiques prosodiques dépend de leur fonction. La narration, la réponse, le titre remplissent phonologiquement la fonction explicative. L'annonce, la déclaration et la nouvelle remplissent phonologiquement la fonction appellative. L'implication et la déclaration emphatique remplissent la fonction expressive. Cette distinction s'explique au niveau prosodématique par des limites d'une marge de dispersion de zones perceptuelles et acoustiques caractérisant le faisceau variable d'indices des prosodèmes correspondants.

#### Références

- Trubetzkoy, N.S. (1960): Osnovy fonologii, Moscou.  
 Jakobson, R., G.M. Fant, M. Halle (1952): Preliminaries to Speech Analysis, Cambridge, Mass., MIT Press.

## THE PATTERNS OF SILENCE: PERFORMANCE STRUCTURES IN SENTENCE PRODUCTIONS

Harlan Lane, François Grosjean and Lysianne Grosjean, Northeastern University, Boston, MA

The pauses produced by speakers while reading familiar material were used to obtain hierarchical sentence structures. Identical structures were obtained from parsing, indicating that the performance structures of sentences are not task specific. The linguistic surface structure of a sentence is a good predictor of the pause durations. However, speakers also revealed a tendency to place pauses between segments of equal length. A simple cyclical model combining, for each pause location, an index of linguistic complexity and a measure of the distance to the midpoint of the segment, accounts for 72% of the pause time variance as opposed to 56% for the linguistic index alone. The generality of the model is shown by its good prediction of the pause durations obtained in unrelated studies in English and American Sign Language.

## PROSODIC LENGTH IN WEST GERMANIC AND SCANDINAVIAN

Anatoly Liberman, Departments of German and Scandinavian, University of Minnesota, Minneapolis, Minn., U. S. A.

There are two main correlations of quantity in modern Germanic languages. One holds sway in West Germanic and to a certain extent in Danish and can be exemplified by such forms as Engl. pulling: pooling, Dan. bære: tælle. The other is represented by Swedish, Norwegian, Icelandic, and Faroese: cf. Swed. pila: pilla, Icel. vina: vinna. The West Germanic type is covered by the concept of Silbenschnittkorrelation (correlation of syllable cut). The prevailing Scandinavian type conforms to the law of syllable length:  $\bar{V}C$  vs.  $V\bar{C}$ . In the forms pulling, tælle, contrary to the forms pooling, bære, the point of syllable division lies within the intervocalic consonants. The same is true of pilla, vinna as opposed to pila, vina. Since phonemes cannot be cut by any linguistic boundaries (by definition), two solutions are possible: either pulling, tælle, pilla have clusters of identical consonants between vowels (i.e. |l+l|) or they lack the point of syllable division altogether. For Swedish, Norwegian, Icelandic and Faroese, the first solution is correct, because in them the complex of the pilla type can conceal a point of word division -pil la-. For English, German, Dutch and Danish, the second solution is only possible, for pulling and the like cannot be taken for a sequence of words the first of which begins with and the second ends in |l| (pulling and 'pul Ling are not homonyms). For all the Germanic languages, length is prosodic, because it is inevitably described in terms of syllables or whole words, but only the Swedish type has geminates. When setting up prosodic length for Germanic languages, one should avoid operating with such criteria as parallelism in the number of long/short phonemes, for the binary division of phonemes according to some feature can have a purely phonematic value (cf. palatalization in Russian or voiced/voiceless obstruents in very many languages).



## A STUDY OF TONE-SANDHI IN STANDARD CHINESE WITH COMPUTER

M.C. Lin, L.H. Lin, G.R. Xia and Y.S. Cao, Institute of Linguistics, Chinese Academy of Social Science

This paper presents the results of measurements of fundamental frequency in Standard Chinese bisyllabic words on a digital computer with the clipping autocorrelation and simplified inverse-filtering techniques. 10kHz sampling rate is used in the former method, and 2kHz in the latter. For that of 2kHz sampling rate, a "real formant" calculating formula is applied to the interpolating compensation in order to obtain the weighting value of fundamental frequency.

142 bisyllabic words of all tone combinations (including the words of "yi" (one), "qi" (seven), "ba" (eight), "bu" (not)) were pronounced by speaker A, while 16 bisyllabic words were pronounced by 3 males and 3 females, respectively.

Experimental results show that when a 1st tone in SC is before or after any other tone, it is always pronounced as high-level although it is generally slightly lowered when placed on the second syllable of a word. The pitch pattern of the 2nd tone is mainly high-rising, but it may be high-falling-rising. However, it is always acceptable to pronounce the 2nd tone as high-rising. A 3rd tone before or after a 1st tone, 2nd tone, and 4th tone is low-falling or low-falling-rising. In case a tone 3 is combined with another tone 3, the first one is high-rising or high-falling-rising. The 4th tone is high-falling. In case two syllables with the 4th tone are put together, the first one does not fall as much as the second one.

The tone alterations of "yi", "qi", "ba" or "bu" are specific for these words and will be discussed in this paper.

The absolute level of pitch may be different for different speakers. Even for the same person, the pitch level may vary, but, in general, the relative pattern of pitch is about the same.

## DIMENSIONS OF TONE SYSTEMS

Ian Maddieson, Phonetics Laboratory, Department of Linguistics, University of California, Los Angeles, CA 90024, U.S.A.

This paper describes patterns in tone systems in terms of an understanding of the relative importance of the dimensions along which tones may contrast, and explains a marked typological dissimilarity between tone and vowel systems as resulting from the different kinds of dimensionalities that underlie tone and vowel systems. Reliable data on tone systems has been assembled through a survey of over 300 tone languages. This survey shows that 2-tone systems are the most frequent. Each added tone reduces the frequency of occurrence. While 2- and 3-tone systems generally have only level tones, both level and contour tones are commonly included in 4-tone systems. 5-tone systems generally include level, rising and falling tones. Contours moving in the same direction but differing in the amount of their pitch change are typically found only in larger tone inventories. Thus, the smaller and more common inventories exploit only contrasts of pitch level, larger inventories add contrasts along a dimension of pitch movement, and the most elaborate and least common inventories also use contrasts of amount of pitch change.

The ranking of these 3 dimensions corresponds with the ranking of the cognate dimensions of average pitch, direction and slope found by Gandour and Harshman (1978) in a study using multidimensional scaling techniques to determine the perceptual dimensions distinguishing an inventory of 9 tone shapes. In this case the ranking implies, roughly, that subjects relied most on average pitch to discriminate between tones, then next they relied on direction, and so on. The correspondence suggests that tone inventories are elaborated by recruiting progressively less salient perceptual dimensions. In contrast to tone dimensions, perceptual dimensions of vowel quality are not ranked in a hierarchical fashion (Terbeek, 1977). However, vowel quality inventories almost invariably contain multiple terms (most frequently 5). Whereas vowel systems seem to be inherently multidimensional, tone systems only become so when they become elaborated.

#### References

- Gandour, J.T. and R.A. Harshman (1978): "Cross-language differences in tone perception: a multidimensional investigation", L&S 21, 1-33.
- Terbeek, D. (1977): "A cross-language multidimensional scaling study of vowel perception", UCLA Working Papers in Phonetics 37.

## AN EXPERIMENTAL STUDY OF TORONTO ENGLISH SENTENCE INTONATION

Ph. Martin, Experimental Phonetics Laboratory, University of Toronto, Canada

Pairs such as English teacher (N + N) "a teacher of English" and English teacher (Adj + N) "a teacher from England" can be differentiated by their intonation patterns, the former bearing a falling melodic contour on the stressed syllable of its first element, the latter showing a general falling contour.

Applying this mechanism to other syntactic categories and to more complex structures, a relatively simple intonation grammar can be built. This grammar generates from the sentence syntactic structure and the type of grammatical categories involved, a sequence of melodic contours, described in terms of features of slope, height and duration, and located on the (primary) stressed syllables of each unit.

Since this theory was essentially developed from British English data, an attempt has been made here to check its validity for Toronto English sentences. Thirty declarative sentences containing from 2 to 9 minimal stressable units were read by 5 Toronto born speakers. Most of the syntactic structures involved were of the type NP + (VP + NP) with different expansions of the verb and noun phrases.

The instrumental analysis of the recorded sentences gave different sequences of melodic contours which were compared to the sequences predicted by the theory. It was found that the experimental results agree with the theoretical sequences in 93% of the cases. The results were particularly satisfactory for the contrast rising-falling related to couples of categories such as Adj + N, N's + N, Adv + V, as opposed to groups such as V + N, N and N, etc., bearing falling contours on their stressed syllables.

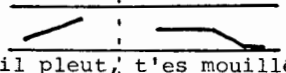
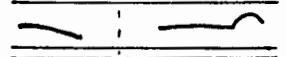
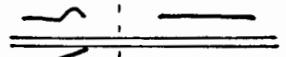
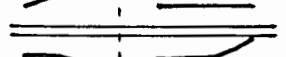
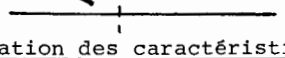
ASYNDETE ET INTONATION EN FRANCAIS.

Alain Nicaise, D.R.L. Université Paris VII et Université Paris XII

A propos du rôle que peut jouer l'intonation dans la mise en relation de deux propositions, cette communication va esquisser une théorie de la représentation des unités prosodiques et formuler des hypothèses sur leurs rapports avec structure syntaxique et intonation.

Les faits

Après une étude acoustique (à l'aide d'un laryngographe) et des tests d'interprétation effectués sur un corpus enregistré de couples de propositions, j'ai dégagé des groupes d'interprétation et analysé la courbe mélodique qui les caractérise. Les résultats exposés à l'aide du couple: "Il pleut, tu es mouillé" sont résumés dans le tableau ci dessous:

Groupe	Schéma mélodique	Interprétation dominante
Groupe 1		"S'il pleut, tu es mouillé"
Groupe 2		"S'il pleut, tu es forcément mouillé"
Groupe 3		"Il pleut, puisque tu es mouillé"
Groupe 4		"Est-ce qu'il pleut? puisque tu es mouillé"
Groupe 5		"S'il pleut, est-ce que tu es mouillé?"

Interprétation des caractéristiques mélodiques

Tous ces schémas peuvent être analysés comme suit: une des propositions (la première ou la deuxième) reçoit une mélodie empruntée à un lexique des mélodies du français, et la courbe intonative que reçoit l'autre proposition est conditionnée par cette mélodie: on peut la dériver à l'aide de règles.

Cette structure mélodique ne reflète pas une structuration syntaxique des énoncés étudiés mais peut être mise en rapport avec une structuration "rythmique" (au sens de M. Liberman, 1975) qui à son tour permet d'expliquer un caractère commun à tous les types d'interprétation de ces couples de propositions: une des propositions sert de repère situationnel à l'autre (elle est le cadre dans lequel l'autre est assertée - ou mise en question).

Bibliographie

Liberman, M. (1975): The Intonational System of English, PhD Dissertation, M.I.T. Unpublished.

## FUNCTIONS OF STRESS AND SEMANTIC STRUCTURE OF UTTERANCE

Tatjana M. Nikolaeva, Institute of Slavistics and Balkanistics  
AN SSSR, Moscow, USSR

1) The present state of linguistics is characterized by the growing interest in stress (prominence) as a main index of semantico-syntactic differences. See, for example, in Russian: jeshche + X ( $X_n$  adds to  $X_{n-1}$ ): jeshche stakan chaju? (Another cup of tea?).

jeshche + X (X adds to Y): jeshche stakan chaju? (And a cup of tea now?).

Odin + X (a certain X); odin + X (one X); odin + X (only X); odin + X (one and only one X); odin + X (the same X); odin + X (there is only X and nothing more).

Tol'ko on ne byl v Pariže (But he wasn't in Paris, you are wrong);

Tol'ko on ne byl v Pariže (Everybody was in Paris, but he wasn't).

2) However, the present state of phonetic and prosodic theory does not correspond to requirements of general linguistics. Namely, the intonation theory only assumes the presence of one main stress (the nuclear) and of emphatic (contrastive) stresses. The semantic interpretation of stress is not yet elaborated, and the stress usually is tied only with the notion of FSP (Actual sentence division). The following problems are not yet solved: 1) How many types of utterance stresses exist on the phonetico-prosodic level? 2) How many semantic categories correspond to these types? 3) What are these categories?

According to our concrete investigations, there are a minimum of 5 types of stress phonetically. They can sometimes co-exist in one utterance (speech unit). These types correspond to the following content components of an utterance: 1) the differentiation of actants; 2) presupposition and assertion; 3) the concrete situation; 4) the connection with text; FSP; 5) the category of definiteness.

## ESSAI D'AUTOMATISATION DE L'ANALYSE PROSODIQUE DU FRANCAIS

Y. Nishinuma et M. Rossi, Institut de Phonétique, Aix-en-Provence, Laboratoire Associé au C.N.R.S., n° 261.

Nous développons une méthode pour l'analyse pluriparamétrique et automatique de la prosodie du français qui se présente comme un modèle d'interprétation perceptuelle. Le programme contient un certain nombre de procédures destinées à extraire, des données brutes, la forme dépouillée de l'intonation.

Organisation du système pour le traitement

Le traitement comprend deux phases 1) l'acquisition des données qui s'effectue automatiquement, 2) le traitement prosodique. Celui-ci comprend un module de gestion, des modules spécifiques pour les conversions perceptuelles et des utilitaires pour le calcul statistique. Dans une 1ère étape, les 3 paramètres acoustiques subissent une série de corrections en fonction des caractéristiques intrinsèques, du contexte et du mode de perception des contours.

Une première stylisation est effectuée à partir de ces résultats. La  $F_0$ , l'intensité et les niveaux intonatifs sont donnés tous les 10 ms en valeurs brutes et normalisées. Dans une 2ème étape, les valeurs des paramètres sont corrigées sur la base des relations syntagmatiques. Les résultats acquis à cette étape sont donnés sous forme trichotomique ; on indique également le degré d'appartenance à l'état déterminé par les calculs. On réduit ensuite ces données en notation binaire en affectant à chaque voyelle l'une des polarités (longue/brève, forte/faible, haute/basse) et on procède à la restylisation graphique des données.

Conclusion

La première version de ce modèle, appliquée au japonais donnait un score de 92% dans la reconnaissance de l'accent. Les premiers résultats obtenus pour le français dans la reconnaissance des unités intonatives permettent d'espérer un score du même ordre sur un corpus étendu. L'exploitation du modèle se révèle utile pour la formalisation grammaticale de la phrase, la synthèse de la parole et la reconnaissance automatique.

## OBSERVATIONS ON RHYTHMICAL UNITS WITH ANACRUSES IN CZECH

J. Ondráčková, Laboratory of Phonetics, Institute for the Czech Language, Czechoslovak Academy of Sciences, Prague, Czechoslovakia

The purpose of this paper is to show the character of both the anacrusis themselves and of rhythmical units with anacrusis, respectively. The material investigated was Czech prose (a language with phonological quantity and fixed stress position) analyzed on the basis of perceptual tests.

Subjects

Monosyllabic anacrusis were examined from the standpoint of grammatical categories. Special attention was devoted to monosyllabic rhythmical units with an anacrusis and to polysyllabic rhythmical units with the anacrusis beginning with a monosyllabic word.

Conclusion

Czech as it is spoken today (and even in the interpretation of the written context) manifests the tendency of using words of all grammatical categories in the function of anacrusis. A great prevalence of conjunctions is to be found in the function of anacrusis.

## THE NEWSREADER'S HIGH FALL

Janina Ozga, Institute of English, Jagiellonian University,  
Kraków, Poland

This paper examines the use of high-falling tone in one variety of Polish, i.e. the language of radio and TV news broadcasts. "High Fall" is treated as a cover term for a set of combinations of falling tone (´) with features from the simple and complex pitch-range systems (terminology and transcriptions of Crystal 1969 are employed). In news-reading, these combinations are always associated with the nuclei of sentence-final tone units.

Until recently, High Fall was hardly ever used in news-reading: its usage was confined to exclamations, commands and statements involving strong emphasis and contrast. The predominating nuclear tone associated with sentence-terminal contours in news-reading was a Low Fall, characteristic of declarative sentences and enhancing the impression of "objective reporting". At present the traditional and the new modes of news reporting exist side by side but they are mutually exclusive in the sense that they are not used as interchangeable variants by any single newsreader.

Possible sources of the newsreader's High Fall are discussed and the explanation which appears to be most convincing, although far-fetched at first sight, is that the intonation is a borrowing from English (e.g. BBC broadcasts), where the High Fall is the predominating nuclear tone associated with the language variety in question. Arguments which support this explanation come from other types of radio and TV language, e.g. the announcements of Polish disc jockeys, news headlines in music programmes, jingles and commercials; BBC and Radio Luxembourg influences in the prosodic stratum are clearly detected.

Reference

Crystal, D. (1969): Prosodic systems and intonation in English,  
Cambridge: University Press.



## DAS PROBLEM DES ANSCHLUSSES IN DEN GERMANISCHEN SPRACHEN

R. K. Potapova, das Moskauer Institut für Fremdsprachen

Zur Zeit gibt es noch keine klare Vorstellung über den Charakter des Anschlusses in den germanischen Sprachen. Die Hauptaufgabe der vorliegenden Arbeit war es, perzeptorische und akustische Korrelate des Anschlusses in den germanischen Sprachen zu finden. Das Ad hoc-Material bestand aus einsilbigen und zweisilbigen Wörtern mit den Lautstrukturen KVK und KV:K im Deutschen, Englischen, Schwedischen, Dänischen und Holländischen. Der Untersuchung lag eine komplexe Methodik zugrunde, die die auditive Segmentierung, Spektralanalyse, die Analyse der Dauer und des Intensitätsverlaufs umfasste.

Die Ergebnisse der Untersuchung berechtigen zu folgenden Schlussfolgerungen:

1. Die auditiven Segmente (ihre Zahl, Reihenfolge und spezifische Merkmale) unterscheiden sich bei langen und kurzen Vokalen und lassen eine Gegenüberstellung nach folgendem Prinzip zu: von geschlossenen zu offenen Segmenten bei langen Vokalen und von offenen zu geschlossenen Segmenten bei kurzen Vokalen.
2. Für kurze Vokale ist die Zentralisierung der gesamten qualitativen Stabilität kennzeichnend, die in der Regel zeitlich mit der Lokalisierung des Kernlautsegments zusammenfällt. Für lange Vokale ist eine frühere qualitative Gesamtstabilisierung und eine spätere Lokalisierung des Kernlautsegments kennzeichnend.
3. Kurze Vokale haben in der Regel einen Gipfelpunkt des Intensitätsverlaufs. Die Zahl der Gipfelpunkte des Intensitätsverlaufs bei langen Vokalen kann infolge der Reartikulation ansteigen. Beim Vergleich der Intensitätskurven wurde festgestellt, dass der Intensitätsverlauf in der Endphase bei langen und kurzen Vokalen im wesentlichen gleich ist.

Die Untersuchung lässt darauf schliessen, dass es sich nicht um verschiedene Typen, sondern grundsätzlich um einen Typ des Anschlusses in den KVK und KV:K-Strukturen handelt. Die kurzen und langen Vokale unterscheiden sich in ihrem Verlauf sowohl auditiv als akustisch, die Endphase des Anschlusses ausgenommen, die in allen untersuchten Parametern im wesentlichen gleich ist.

## CONTOURS MELODIQUES SYLLABIQUES ET TONS A NIVEAUX EN TERRASSES

Annie Rialland, Institut de linguistique, Université de Paris V,  
12 rue Cujas, 75005 Paris

Le système tonal du gulmancema (nom vernaculaire du gourma, langue Gur, 400000 locuteurs) n'a pas fait l'objet d'études jusqu'à maintenant. Il s'agit, en fait, d'un système à trois tons "à niveaux en terrasses", donc très proche de celui du yoruba.

L'étude des réalisations des tons à l'aide de mingogrammes nous a permis d'arriver à la conclusion que chaque ton se différencie des autres par deux traits: une hauteur et un contour mélodique.

Ceci est net, à la fois dans les monosyllabes où les trois tons présentent un contour mélodique spécifique et dans la chaîne où ces courbes se retrouvent. Ces deux traits habituellement associés peuvent être dissociés.

Dans certains contextes, certains tons présentent une courbe mélodique sans modulation. Ils ne sont plus, alors, identifiés que grâce à leur rapport de hauteur avec les syllabes adjacentes. C'est le trait "hauteur" qui prend seul en charge la fonction distinctive.

Mais, le trait "contour" peut également être séparé du trait "hauteur". Ainsi, un ton bas relevé (le relèvement de la première syllabe du nom est la marque de la possession et de la détermination par une proposition) continue à se distinguer d'un ton haut, grâce à son profil mélodique descendant. Il en est de même pour le ton moyen qui reste mélodiquement plat.

En cas d'élimination de voyelle, le ton se maintient à travers le seul trait de hauteur. Il n'est plus représenté que par la hauteur du point de départ de la courbe mélodique du ton suivant.

Deux traits caractérisent donc chaque ton, et il faut insister sur la fréquence de la présence du trait "contour" même si on pose que celui-ci est redondant. Les descriptions du système yoruba mentionnent des faits proches quoique ne concernant pas tous les tons.

L'étude d'autres langues permettrait peut-être de mettre en évidence une relation entre tons "à niveaux en terrasses" et tons caractérisés par les traits de hauteur et de contour.

En effet, les hauteurs et les intervalles sont variables au cours de la phrase et ne peuvent pas jouer le même rôle pour la discrimination des tons que dans des langues à niveaux discrets. D'où peut-être l'importance des contours pour faciliter la reconnaissance des tons.

## AUTOMATIC DETECTION OF PROMINENCE IN THE DUTCH LANGUAGE

A.C.M. Rietveld and L. Boves, Institute of Phonetics, Katholieke Universiteit Nijmegen, Erasmuslaan 40, Nijmegen, The Netherlands

The procedures which will be described in this contribution aim at an automatic detection of prominence (sentence-stress) in Dutch. We regard this as the first step towards an automatic transcription of the intonation of this language. The physical correlates of prominence are essentially three pitch-movements with specific characteristics as has been shown by 't Hart and Collier (1975). The procedures which are going to be described have to detect those fragments of the Fo-curve which can be regarded as realizations of the prominence-lending movements mentioned above.

After the Fo-variations have been measured with an analog pitchmeter, the resulting curve is converted into a digital signal for further processing. First a correcting program smoothes out the comparatively irregular curve and rejects outliers. Then an approximation procedure transforms the curve into a series of straight lines by applying a Least Squares criterion, together with an "error" criterion in order to interrupt the approximation of a segment if the error exceeds a certain value.

A labeling program labels the resulting straight lines and tries to combine adjacent segments with similar characteristics into a smaller number of lines which are then given the same label. This program is partly based on the principles of linguistic pattern recognition.

The detection of prominence is carried out on the basis of three sources of information: the above mentioned labels, the syllable-structure, and the amplitude of the syllabic segments.

Preliminary tests with three short texts resulted in detection-scores of 71%, 75%, and 92%, respectively.

The description of our procedures is completed by some experiments in which relevance and value of the performance-criterion - stress-judgments - are examined. These experiments, which involved the manipulation of the Fo-contours of utterances, showed that listeners may "switch" from pitch to other acoustic cues when trying to determine prominence in monotonous speech. This result implies that listener-judgments are of limited use for the evaluation of the performance of our detection procedures.

Reference: 't Hart, J. and R. Collier (1975): "Integrating different levels of intonation analysis", JPh 3, 235-255.

## FURTHER OBSERVATIONS ON SECONDARY STRESS IN BRITISH ENGLISH

Alan E. Sharp, Department of Linguistics, University College of North Wales, Bangor, Gwynedd, U.K.

At the last Congress I drew attention, under the heading "The evil 'i': or, shellfish is not 'somewhat shellf'. A pitfall for the unwary in English stress." to the frequent failure of phonetically transcribed texts to distinguish accurately and consistently between 'light' and 'heavy' /i/ (English Pronouncing Dictionary transcription) in places away from the tonic or nuclear stress: in other words, to recognise on this particular vowel quality the incidence of secondary stress. In this paper I examine two other areas in which problems may arise in connexion with secondary stress in British English.

'Light' and 'Heavy' Diphthongs and 'Long Pure' Vowels

In many words, notably in the immediately pre-tonic position, 'light' versions of the named categories may occur which are, on a lexically selective basis, distinct from the traditional weak vowels. Psychiatrist may show this phenomenon, psychology not.

Isochronicity

Where it is possible to establish a fully satisfactory 'foot' or 'stress bar' internal analysis may make it possible to distinguish minor prominences from surrounding weak syllables and to refer these prominences to secondary stress. The status, however, of isochronicity is suspect and in the absence of additional criteria grave difficulties of identification persist.

Both problem areas reflect the need for an exhaustive analysis of the temporal organization of The English utterance.

## NOTES ON MELODIC HOMONYMY IN STANDARD RUSSIAN

Jan Skoumal, Prague, Czechoslovakia

Unlike the type of homonymy discussed by Romportl (1973), which means a coincidence of all manifestations of two different melodemes, many types of homotony observed in Russian seem to be rather a result of overlapping of some manifestations of one melodeme with some manifestations of another melodeme. Analysis shows that, although this overlapping often produces a total melodic identity of two distinct utterances, the underlying melodemes are not neutralized for they preserve their specific features and structures.

The paper discusses certain cases of melodic figures whose interpretation within the framework of Bryzgunova's intonation structures (1963 et al.) causes difficulties, and suggests a re-interpretation to account for their formal and functional properties in a more satisfactory manner.

Conclusions

1. There seems to be a regular correspondence between the melodeme and the physical parameters of its allomels. In Russian, e.g., melodeme is responsible, among other things, for tone properties of the relevant points as well as for their local distribution over the given segment.
2. Homonymy of allomels in Russian is made possible by the fact that relevant points in segments realizing different melodemes may have different locations (e.g., ictic or final); consequently, a stretch of melody which is less relevant for one melodeme may vary in such a way as to coincide with a highly relevant stretch of a different melodeme.
3. Homonymy (homotony) of those allomels which realize different melodemes does not, unlike neutralization, mean a suppression of semantic distinctions between utterances.

References

- Bryzgunova, E.A. (1963): Praktičeskaja fonetika i intonacija ruskogo jazyka, Moskva: Izd. MGU.
- Bryzgunova, E.A. (1977): Zvuki i intonacija ruskoj reči, 3-e izd., Moskva: Russkij jazyk.
- Romportl, M. (1971): "K synonymii a homonymii intonačnich pro- středků", in Slavica Pragensia 13, 209-217.
- Romportl, M. (1973): Studies in Phonetics, Prague: Academia.

## ZUR SEMASIOLOGISCHEN INTERPRETATION DER SPRECHMELODIE

Eberhard Stock, Halle-Neustadt, Deutsche Demokratische Republik

In der Arbeit wird begründet, weshalb die in der Intonationsliteratur über das Deutsche vertretene Auffassung über das Vorhandensein von lediglich 3 syntaktisch-relevanten prosodischen Zeichen mit ein-eindeutiger Form-Bedeutungs-Relation der kommunikativen Realität nicht angemessen ist.

Die Überprüfung des Inventars der melodischen Formen ergibt, dass die prosodischen Zeichen grundsätzlich durch andere meist lexikalische Mittel ersetzbar sind, bzw. dass sie mit bestimmten lexikalischen Mitteln kombiniert werden müssen, um unter den jeweiligen Kommunikationsbedingungen die intendierte sprachliche Bedeutung signalisieren zu können. Ihre funktionelle Belastung bzw. Belastbarkeit ist daher geringer als von vielen Phonetikern angenommen. Darüber hinaus muss aus einigen empirischen Untersuchungen geschlossen werden, dass die Endphasenmelodie in bestimmten Arten von Kommunikationsereignissen auch als phonostilistisches Mittel genutzt werden kann und dann syntaktisch funktionell nicht belastet wird.

Hieraus resultiert, dass die letzte und entscheidende Determination für die Melodisierung beim Sprechen nicht in der Syntax und auch nicht in der semantischen Komponente der Grammatik gesucht werden kann, sondern in der kommunikativ pragmatischen Orientierung des Sprechers, die der syntaktischen Motivierung der Sprechmelodie in vielen Fällen übergeordnet ist. Diese Tatsache muss in Generierungsmodellen berücksichtigt werden.

SOME EXPERIMENTS IN THE DIGITAL EXTRACTION OF AMERICAN  
INTONATION PATTERNS

Yukio Takefuta, Chiba University, Chiba, Japan and Osaka  
University, Osaka, Japan

Many models of intonation patterns for American English have been proposed by linguists and phoneticians. However, none of them, including both auditory-analysis models and instrumental-analysis models, seems to be accepted as valid and reliable. We believe that one of the best approaches to the study of intonation is the one in which the instrumentally extracted physical data (the frequency contour) are processed until they match the auditory recognition of intonation contrast. A number of data processing techniques (normalization, transformation, and feature extraction) were tested and introduced in this study to increase the validity of the patterns to be obtained by the instrumental technique. The principle of "relevancy" or "distinctiveness" in linguistic signals and the computation of the index of signal detectability ( $d'$ ) were used as the criteria to measure and compare the validities of the developed computer algorithms.

The computer programs developed in this study were tested at each step of the data processing using three sets of utterances, and also in an integrated set of a simulation program developed for determining intonation patterns. The digital technique was found to be useful to determine intonation patterns of American English for a theoretical study of the intonation system, and also for a practical application of building a teaching machine which can help foreign students or persons with speech problems learn American intonation effectively.

ATTITUDINALLY DISTINCTIVE FUNCTION OF SOME PRE-TONIC STRESS-AND-PITCH PATTERNS IN ENGLISH IN REFERENCE TO THEIR PHONO-STYLISTIC USAGE

I.S. Tikhonova, Moscow State Pedagogical Institute, English Department (USSR)

The attention of phoneticians has recently been drawn to the rapid development of phonostylistics and a completely new approach has arisen to the investigation of phonetic phenomena.

In syntactic phonetics the prosodic phenomena are now analysed inseparably from their realization in certain speech styles taking into account the purposes of communication.

The pre-tonic stress and pitch sections of a phrase have been thoroughly studied up till now together with the terminal tones. There appeared recently a number of books and articles both at home and abroad in which different tonetic and stress-and-pitch patterns are presented as a system and in correlation with certain communicative types.

Some scientists, however, have started investigating pre-nuclear sections of the pitch, irrespective of terminal tones.

Nevertheless, not much is known yet about what functions are fulfilled exclusively by heads. It is generally acknowledged that heads are very important sections of pitch patterns. But it has not been proved yet that their attitudinal function is dependent on speech situations, different types and styles of speech.

Having analyzed a great number of sound texts and sources on the subject, we attempted to single out structural types of descending heads in English: the falling head, the stepping head, the scandent head, and the sliding head. The first three differ in the direction of unaccented syllables in the head; the last - by the sliding down variations on the accented syllables.

In this paper we managed to prove that each type of descending heads in English is distinctively different and has its own intonological status. They express different attitudes of a speaker towards the utterance or reality, but their attitudinally distinctive function varies in different styles of speech. So their modality, and attitudinal difference should be studied as applied to intonation styles. The problem of head tonemes is not fully solved yet, and we are hopeful that in this paper one more step towards its solution is made.



## L'ETUDE DE L'INTONATION ET LA THEORIE DU TEXTE

I.G. Torsouéva, Institut des Langues Etrangères, Moscou, URSS

L'étude de l'intonation en tant que partie intégrante du texte peut jeter une lumière sur des problèmes irrésolus de la théorie du texte. Parmi ces derniers les plus importants sont les suivants: recherches des indices formels de la clôture et de l'autonomie du texte, principes du démembrement, volume et caractéristiques de ses composants, moyens d'expression des fins communicatifs du texte, procédés de la syntaxe transphrastique. La solution de ces problèmes proposée par la grammaire du discours est insuffisante; l'étude du texte exige une analyse complexe dans le cadre de laquelle l'intonation joue un rôle considérable.

L'intonation organise le texte de manière différente dans la lecture et dans la conversation spontanée. Notre analyse concerne surtout la lecture. Dans ce cas l'intonation remplit des fonctions primordiales: structuration du texte comme un tout, définition du style et de la place du texte en question dans l'ensemble des autres, démembrement du texte en ses composants, liaison entre ces composants, établissement des rapports transphrastiques, influence sur l'allocutaire.

Les procédés intonatifs des rapports transphrastiques sont: corrélation des niveaux mélodiques à la jonction de deux phrases, parallélisme des constructions intonatives, contraste des constructions intonatives, organisation rythmique, pauses, etc.

L'étude de différentes formes du texte prouve qu'il n'y a pas de correspondance biunivoque entre la liaison sémantique et la liaison intonative de deux phrases voisines. La connexion intonative peut s'avérer comme moyen de compensation en l'absence de relations sémantiques. Les résultats de notre expérience seront présentés dans notre exposé.

## JOINTURES EN FRANCAIS ET STRUCTURE PROSODIQUE

Jacqueline Vaissière, Centre National d'Etudes des Télécommunications, 22300 Lannion, France

Cette communication présente une méthode de dérivation de la structure prosodique d'une phrase à partir de l'analyse des pauses et de la courbe du fondamental (Fo) illustrée pour la langue française. Dans une première partie, la méthode d'interprétation de la courbe de Fo est rappelée: (1) Quantification de la courbe par valeurs-cibles sur les voyelles, (2) Interprétation en termes de mouvements (tels que R-rise, L-lowering, etc...), (3) Interprétation des mouvements en termes d'attributs qui sont des mouvements identifiés selon leur position dans le mot, (4) Interprétation des suites de mouvements en patterns. La conclusion est qu'un mot lexical peut être prononcé de 5 façons différentes: schéma montant, descendant, à pic, plat, ou parenthèse. La seconde partie traite de l'utilisation faite par les locuteurs de la possibilité de regrouper plusieurs mots lexicaux en un seul pattern (1), et celle de prononcer un même mot de plusieurs façons (2). On verra alors, qu'avec cette troisième possibilité qu'a le locuteur de placer les pauses où il veut, les combinaisons de patterns permettent de créer différents degrés de jointure, que nous avons classés dans l'ordre suivant (par ordre d'importance pressentie):

1. Pattern descendant + Pause (descente finale, avec maximum local de Fo sur la dernière syllabe de l'avant dernier mot). (cf: fin de phrase). -
2. Pattern montant + Pause (Montée de continuation avec minimum local sur l'avant dernière syllabe). -
3. Succession de trois mouvements de sens contraire R,L,R tel qu'entre un pattern à pic et un autre pattern (sauf parenthèse). -
4. Les combinaisons de patterns créant la succession de deux mouvements de sens contraire: L + R (pattern plat suivi d'un autre pattern sauf parenthèse), ou R + L (pattern à pic suivi du pattern parenthèse). -
5. Les combinaisons ne faisant intervenir qu'un seul mouvement: R (pattern parenthèse ou descendant suivi d'un autre pattern sauf parenthèse), ou L (pattern plat suivi du pattern parenthèse). -
6. Les jointures réalisées sans mouvement R ou L.

Grâce à ce système, il est possible de dériver la structure prosodique réalisée dans les phrases par les locuteurs, et d'avoir ainsi à sa disposition une base commode pour comparer les choix des différents locuteurs ou de comparer la structure prosodique et la structure syntaxique.

## THE ACCENTUAL AND MELODICAL STRUCTURE OF STYLISTICALLY HETEROGENEOUS TEXTS. AN INVESTIGATION BASED ON ENGLISH LANGUAGE MATERIAL

Galina M. Vishnyevskaya, Ivanovo State University, Ivanovo, USSR

This paper presents a short account of an investigation aimed at discovering certain peculiarities of the accentual and melodic organization of stylistically heterogeneous recorded English texts (transposed from the written form of English into their sound form).

An auditory and linguistic analysis of experimental texts (excerpts from a fairy-tale, a novel, a lecture, some humorous stories, some conversations, presenting different functional styles) shows that the structural elements in all of them have specific quantitative and qualitative features.

Their structural differences display themselves, for example, in the number of syntagms (or breath-groups) with reference to the general number of sentences in the text, in the frequency of occurrence of the main types of nuclear tones of English, and in the frequency counts of the accentual types of the syntagms.

It has been discovered that, alongside with the nuclear tone, the scale (or the head, that part of the accentual structure which extends from the first stressed syllable up to the nucleus) has a very high frequency of occurrence: 68% of all the syntagms comprising the scale.<sup>1</sup>

#### Conclusion

A preliminary survey of the results leads us to the conclusion that the stylistic heterogeneity of the texts determines their peculiar accentual and melodic organization, which requires further investigation and description.

#### Reference

Crystal, D.D. (1969): Prosodic systems and intonation in English, Cambridge.

---

(1) These data correspond to those presented by the English phonetician D. Crystal (1969).

RULES OF INTERACTION BETWEEN SEGMENTAL AND SUPRASEGMENTAL  
FEATURES IN THE ORGANIZATION OF SPOKEN TEXT

L.V. Zlatoustova, Laboratory of Structural and Applied  
Linguistics, Moscow University, USSR

An acoustic and auditive analysis of artistic prose and spontaneous colloquial Russian speech revealed the following regularities:

The most important suprasegmental characteristics of the stressed syllable of a phonetic word are a relative increase in duration, in particular of the vowel, and a close approximation of the formant frequencies of the vowels to their ideal or target values. A Russian phrase consists, on the average, of 2.2 - 3.2 phonetic words, and the stressed syllables show a marked  $F_0$  movement at the beginning and, in particular, at the end of the phrase. Thus, for phrase initial and phrase final phonetic words,  $F_0$  is a third parameter which helps identifying both the placement of stress and the placement of phrase boundaries.

The next level of suprasegmental organization is the phrase level. The partitioning of continuous speech into phrases manifests itself by phrase stresses and pauses. Note that the pause has a linguistic status: logical and emotional emphasis, statements, consequence and result can be signified by filled pauses. A filled pause is manifested in a phrase final vowel or consonant; thus, segmental units have a suprasegmental function in these cases.

Phrase stresses are superimposed on the stressed syllables of phonetic words. According to their function, phrase stresses belong to either of two classes:

The first class is formed by those phrase stresses whose function is to organize the phonetic words in the phrase. Their most important acoustic parameter is  $F_0$ .

The second class comprises emotional and logical stress. The function of these types of stress is to emphasize certain parts of the speech chain. The acoustic characteristics of such types of stress interact to a high degree with those of the segmental units: Emotional stresses are marked by a change in the spectral pattern of the vowels, by an increase in duration of the stressed syllable, etc.

## SOME PHONETIC DIFFERENCES BETWEEN ARABIC AND ENGLISH VOWELS

Mohammad Anani, University of Jordan, Amman

Arabic vowels have not been satisfactorily described yet. The reasons are, in the main, due to "pedagogic" presentation, relegation of important phonetic facts to "irregular" or "stylistic" status and lack of interest in contrastive relationships obtaining between vocalic elements.

It is hoped that this article will give a more orderly presentation of Arabic vowels and take more fully into account important features which, in previous analyses, have been regarded as "redundant".

Differences between the Arabic vocalic system and the English vocalic system are mentioned.

## ADAPTATION PHONETIQUE D'UNE METHODE DE REEDUCATION DE L'APHASIE

D. Autesserre, N. Scotto Di Carlo, M.C. Hazaël-Massieux,  
 Institut de Phonétique d'Aix-en-Provence

En 1973, Albert, Sparks & Helm ont mis au point une nouvelle technique de rééducation de la parole chez l'aphasique - Melodic Intonation Therapy - (MIT) qui s'adresse à des cas sévères pour lesquels la démutisation résiste aux méthodes traditionnelles. L'originalité de cette technique réside dans l'introduction d'une étape intermédiaire pendant laquelle on fait répéter au malade des phrases courtes en voix chantée sur des mélodies arbitrairement choisies. C'est dans une seconde étape seulement que l'on aborde la répétition des phrases parlées proprement dites. Il nous a paru intéressant d'adapter cette démarche rééducative en fournissant au malade des modèles de voix chantée reproduisant exactement l'intonation des phrases parlées qui leur seront présentées par la suite.

Afin d'éviter que le patient ne revienne très vite à une voix inexpressive tendant vers le recto-tono, les phrases parlées à partir desquelles on a dégagé les lignes mélodiques de la voix chantée ont été prononcées avec des contours intonatifs expressifs.

Un soin tout particulier a été accordé à la formation de l'orthophoniste afin que les lignes mélodiques des modèles soumis aux malades tant en ce qui concerne la voix chantée que la voix parlée, ne varient pas d'une séance à l'autre. Cette façon de procéder permet de suivre les performances de l'aphasique et de contrôler expérimentalement l'adéquation des répétitions au modèle proposé. Par ces possibilités de validation de la rééducation et les résultats très encourageants obtenus avec les malades, cette adaptation du MIT apporte une contribution importante à la réadaptation d'aphasiques présentant des troubles sévères et persistants pour lesquels toute autre approche rééducative avait échoué.

#### Références

- Albert, M., R. Sparks et N. Helm (1973): "Melodic intonation therapy for aphasia", Archives of Neurology 29, 130-131.
- Sparks, R., N. Helm et M. Albert (1974): "Aphasia rehabilitation resulting from melodic intonation therapy", Cortex 10, 303-316.

## SOME PROBLEMS OF ALPHABETS AND THE RUSSIAN ORTHOGRAPHY

Uzbek Sh. Baitchura, Leningrad

Problems of Russian orthography (and of other languages in Russia) have been much discussed, thousands of improvements have been suggested lately, and this gives evidence of the unfitness of the Cyrillic alphabet adopted together with the Greek branch of Christianity and ever since being gradually latinized, although the process has not yet been completed.

In the 20es, after the Arabian alphabet of the Moslems of Russia was replaced by the Latin alphabet "to bring these peoples nearer to the higher European culture", although the Arabic script (as other alphabets of Semitic origin) was superior to the Latin and especially to the Cyrillic ones (G. Sharaf et al., 1926), the problem of replacing the Cyrillic alphabet of Russians by the Latin one arose. The idea was supported by V.I. Lenin, by the 1st Minister of Education A.V. Lunačarskij, a.o., a "Subcommittee on Latinization of the Russian Alphabet" was established at the Ministry of Education; its president was Prof. N.F. Yakovlev, and its resolution (1930) stated that the Russian alphabet was "an anachronism", separating us from the West and the East, a means of "russification", of "national oppression" and that "a new alphabet must be adopted... in keeping with the international content of socialistic culture" (i.e. Latin). But Lenin and Lunačarskij died, the project was dropped, even the new national Latin alphabets were replaced by the Russian one without discussions or explanations.

Nowadays, when the Latin alphabet has become the principal means of international contacts, it is time to revive the plan of replacing the Cyrillic alphabet by the Latin one (used also by many Slav peoples, Estonians, Letts, etc.), which is justified from the scientific and practical points of view, as the Latin alphabet is superior to the Cyrillic one in all respects and answers the spirit of our time, consisting in a general trend toward international collaboration, but not in national or governmental isolation or confrontation (including that on the level of alphabets).

References

Vsesojuznyj tjurkologičeskij sjezd. Stenografičeskij otčet. Baku, 1926, 242-260 et passim.

Kul'tura i pis'mennost' Vostoka VI, Baku 1930, 20-43, 208-216.

## A STOCHASTIC MODEL OF PHONEMIC PATTERNS IN SPOKEN ITALIAN

U. Bortolini<sup>\*</sup> - F. Degan<sup>\*</sup> - C. Minnaja<sup>\*</sup> - L.G. Paccagnella<sup>\*,\*</sup>,  
Centro di Studio per le Ricerche di Fonetica (C.N.R.), Padova, Italy  
<sup>\*</sup>Istituto di Matematica Applicata, Università, Padova, Italy

Many situations in automatic speech recognition/understanding require decisions which have often to be made on the basis of incomplete or uncertain information. Stochastic modeling is a flexible general method for handling such situations. It consists of employing a specific probabilistic model which helps in uncertainty or incompleteness of the information. In this paper a specific class of stochastic model is discussed - models based on the theory of Markov processes.

The order of the source has been fixed on the basis of the average length of phonetic words. Transitional probabilities are estimated from the occurrences of sequences of two and three phonemes. Such statistics are evaluated from a corpus of spoken Italian, consisting of 49.533 phonemes, derived from 7.667 phonetic sequences.

The construction of the model is very simple, and it is based on drawing out random numbers. The sequences generated by our model point out phonotactic restrictions which are peculiar of the Italian language. The phonetic and syllabic structure of the sequences obtained reflect with good approximation structures which are the most frequent in the natural language.

First some statistics are exhibited and the proprieties of the general model are discussed; then we considered some examples of situations in automatic speech analysis in which such a model can be applied.

#### References

- Baker, Y.K. (1975): "Stochastic modeling for automatic speech understanding" in Speech Recognition, Reddy D.R. (ed.), 521-542, New York: Academic Press.
- Bortolini, U., Degan, F., Minnaja, C., Paccagnella L., Zilli, G. (1978): "Statistics of spoken Italian" in Proc. Ling. 12.



## SUPRAGLOTTAL AIR PRESSURE VARIATIONS ASSOCIATED WITH CONSONANT COGNATE PAIRS PRODUCED BY DEAF PERSONS

W.S. Brown, Jr. and Sue Newman, IASCP and Department of Speech, University of Florida, Gainesville, Florida 32611

The current research further investigated whether there are consistent patterns within deaf speech (a deaf phonology), utilizing measures of supraglottal air pressure ( $P_{10}$ ); these patterns were compared with those of normal hearing speakers. Specific measures examined were overall peak  $P_{10}$  values, peak  $P_{10}$  variations for voiced versus voiceless distinction, and influence of syllabic position and vowel context for consonant cognate pairs. It was also the purpose of this study to determine the constancy of production of deaf speakers over repeated utterances. The cognate pairs /t/, /d/ and /p/, /b/ were combined with the vowels /i/ and /a/ in vowel-consonant-vowel, consonant-vowel, and vowel-consonant forms; all syllable combinations were produced in the carrier phrase, "Say -- again". In addition, sentences specifically composed of words containing the consonants /p,b,t,d,s/ and /z/ were repeated five times each in succession. Five congenitally deaf children (a puretone average greater than 90 dB HL) with semi-intelligible speech repeated the speech sample.  $P_{10}$  was recorded via a custom fitted air pressure sensing tube molded to fit around the speaker's premaxillary arch. The resultant  $P_{10}$  traces were displayed on an oscillographic recorder. Measurement and analysis of the data indicated similar trends when comparing deaf speech to that of normal hearing speakers, especially the constancy of repeated utterances. These data indicate that deaf speakers exhibit a phonology which, although more inconsistent, is similar to that of normal hearing speakers. These results will be discussed in terms of rehabilitation of deaf communication (verbal).

ETUDE DE LA REGULATION NEURO-MOTRICE DES PARAMETRES PROSODIQUES  
A PARTIR DES PRODUCTIONS DE MALADES NEUROLOGIQUES

C. Chevrie-Muller, N. Cerceau et C. Guidet, Institut National de la Santé et de la Recherche Médicale, U3 - Hôpital de la Salpêtrière, Paris, France

Dans la régulation des phénomènes prosodiques on a à apprécier le rôle joué par les structures corticales ("centres du langage"), mais aussi celui dévolu aux structures sous-corticales qui assurent, notamment, le contrôle de la motricité automatique. On se propose d'aborder, à partir des productions de malades atteints d'affections neurologiques, l'étude de ces mécanismes sous-corticaux.

L'échantillon de malades a été retenu à partir de critères acoustiques et neurologiques. Sur le plan acoustique les sujets devaient présenter des altérations des paramètres: fréquence fondamentale (et sa modulation), intensité (et sa modulation), débit de la parole. Sur le plan neurologique on a tenu à éliminer: 1) les troubles psychiatriques, 2) les aphasies et anarthries par lésions corticales hémisphériques, 3) des difficultés "mécaniques" soit par paralysie des muscles agonistes, soit par hypertonie des antagonistes (la difficulté, ou l'impossibilité, de mobiliser les effecteurs, entraîne des troubles de la réalisation articulaire, le trouble prosodique relève alors d'un mécanisme évident). Sur 1100 enregistrements de malades, 96 ont pu ainsi être retenus.

Des corrélations ont été établies entre les symptômes neurologiques et les symptômes acoustiques. Les enregistrements ont été jugés selon un protocole standard d'écoute et les paramètres acoustiques (fréquence fondamentale, durée) ont été mesurés soit sur enregistrements oscillographiques, soit plus récemment par analyse programmée du signal acoustique par ordinateur (Chevrie-Muller et al., 1973).

On a confirmé le rôle joué par les structures cérébelleuses et extra-pyramidales. Des hypothèses ont été formulées sur le niveau lésionnel de syndromes rares, notamment le dérèglement de la hauteur de la voix (450 à 700 Hz, chez 2 femmes) et l'aprosodie à la suite de comas post-traumatiques.

Référence

Chevrie-Muller, C. et P. Decante (1973): "Etude de la fréquence fondamentale en pathologie", Bull. Audiophonol. (Besançon), 3, 147-194.

CONTACTS DE LANGUES: ETUDE PHONETIQUE D'EWONDOS  
FRANCOPHONES A YAOUNDE

Jean-Roland Deltel, Faculté des Lettres et Sciences Humaines  
de Yaoundé, Cameroun

La comparaison des systèmes phonétiques français et ewondo chez un même locuteur ewondo francophone à Yaoundé au Cameroun éclaire le problème des interférences entre langue première et langue seconde en général et celui particulier du français en Afrique.

Des ewondos francophones ayant une bonne connaissance du français, niveau primaire et secondaire ont été invités à prononcer une série de phrases en français et en ewondo analysées ensuite au laboratoire et contrôlées par des enregistrements des mêmes phrases prononcées par des locuteurs ewondophones très peu francophones d'une part et des francophones non ewondophones d'autre part.

Le français parlé par un ewondo francophone révèle par rapport au français standard les caractéristiques suivantes: une compression et une centralisation du système vocalique, des phénomènes de glotalisation, de palatalisation et de labiovélarisation consonantiques, un affaiblissement des occlusives et une réduction des groupes consonantiques, un renforcement surprenant des voyelles nasales et une nasalisation parasite des voyelles orales au voisinage des occlusives sourdes. Ces caractéristiques s'expliquent aisément par l'influence de l'ewondo également analysé et peuvent être mises en relation avec le système phonologique et phonétique de la langue première.

Cependant d'autres phénomènes se révèlent encore plus caractéristiques: modification des schémas intonatifs, des groupes rythmiques, bouleversement du schéma accentuel et de la réalisation de l'accent, et surtout un allongement vocalique général hyperréalisé. La relation entre ces phénomènes et les structures de l'ewondo ne semble pas aussi aisément démontrable que pour les caractéristiques phonématiques précédentes, particulièrement en ce qui concerne la réalisation de l'accent et l'allongement vocalique.

La réalisation d'une langue seconde comporterait, outre des interférences phonématiques purement linguistiques, tout un jeu de caractéristiques prosodiques relativement indépendantes de la langue première peut-être d'ordre psycholinguistique qui expliqueraient en ce qui concerne le français l'existence à la fois d'un "accent africain" et de variétés phonétiques très diversifiées.

FIRST LANGUAGE PHONETIC PERSEVERATION: A THEORETICAL EXPLANATION  
Don George, Univ. Southern Mississippi, Hattiesburg, Miss. U.S.A.

The difficulty often experienced in acquiring an accurate production of the sounds of a second language or dialect is familiar to all. Scientific research into the structure of the brain and the function of its various parts has thrown considerable light on much that until recently was little known. Assuming that total understanding of the entire complex relationships involved may never be realized, we may still attempt inferentially to reach a possible explanation of the difficulty generally experienced in acquiring a second language phonetic system.

We shall assume, for purposes of this paper, that the difference in phenomena observed under controlled laboratory conditions and phenomena found in the real-life operations of human language are differences of complexity and not differences of kind.

The articulatory muscle movements required for any language are stabilized whenever it is found that the sounds being produced are acceptable to others speaking the language. Since the phonological system of any language is a closed system, while the number of possible utterances in the language is an open system, kinesthetic memory of the phonology is more strongly reinforced than any particular syntactic combination. It will be shown that internal feedback from kinesthetic memory overrides any differing auditory stimulus. The speaker of a second language "feels" that the sounds he is producing are the same as those made by the native speaker, particularly when the second language has been internalized to a high degree of fluency. Even when aware of the difference he often finds it difficult to adjust his own articulatory musculature to the difference.

This paper proposes to provide a theoretical base by which teachers of second languages may attack the problem of second language phonology, and from which further research into the application and validity may be undertaken.

## PERCEPTUAL AND ACOUSTIC ANALYSIS OF VOCAL DYSFUNCTION

Britta Hammarberg and Björn Fritzell, Institute of logopedics and phoniatics, Huddinge University Hospital, Jan Gauffin and Johan Sundberg, Speech Transmission Laboratories at the Royal Institute of Technology, Stockholm, Lage Wedin, Institute of Psychology, Stockholm University, Stockholm, Sweden

There is a great need in phoniatic-logopedic diagnosis and treatment for objective criteria of vocal dysfunction. Today voice analysis relies mainly on subjective visual and auditory observations. To make research methods for acoustical voice analysis clinically applicable, a project has been carried out in cooperation between the Institute of logopedics and phoniatics at Huddinge University Hospital and the Speech Transmission Laboratories at the Royal Institute of Technology, Stockholm. Clinically experienced logopedists and phoniaticians evaluate 32 pathological and normal voices in respect to 26 perceptual variables on a 5 point scale concerning voice quality and pitch. A standard text (about 40 sec) is being read by the subjects and the signal is recorded on a two channel tape recorder. The signal comes from a spectacles-worn microphone with a constant mouth-to-microphone distance on one channel and on the other channel from a contact microphone put on the throat below the thyroid cartilage.

The evaluations of the voices are analyzed by factor analysis (Principal Component Analysis). The resulting factors are compared with acoustic measures from mainly three types of analysis: long time average spectrum analysis (LTAS), and distribution analysis of the fundamental frequency, which is performed on the signal from the contact microphone. In order to analyze time bound characteristics of the voice signal a frequency-perturbation measure is also being used.

The results of the perceptual evaluation and of the acoustic measures are being compared by means of multiple regression analysis.

Reference

Fritzell, B., B. Hammarberg, L. Wedin, J. Gauffin and J. Sundberg (1977): "Clinical applications of acoustic voice analysis", Speech Transm. Lab. - Quart. Progr. and Status Rep., Royal Inst. of Techn., Stockholm 2-3, 31-43.

UNTERSUCHUNGEN ZUR PHONEMATISCHEN EINORDNUNG DER ZISCHLAUT-  
STÖRUNGEN

Jürg Hanson, Hals-Nasen-Ohren-Abteilung (mit Phoniatrie) des Kreis-  
krankenhauses, DDR-Eberswalde

Wie aus den meisten Literaturangaben hervorgeht, werden die Zischlaute im allgemeinen den S-Lauten beigeordnet. Das hat zur Folge, dass auch Störungen der Zischlautbildung unter dem Sammelbegriff "Sigmatismen" zusammengefasst und dementsprechend bewertet werden. Krech, v. Essen, Wängler u.a. führen als Begründung dafür an, dass bei der Bildung des normalen [s] und [ʃ] die Artikulationszone praktisch die gleiche sei, das [ç] wird offenbar mangels Einordnungsschwierigkeiten gewöhnlich gesondert erwähnt.

Wir überprüften dazu innerhalb von drei Jahren 275 Patienten einer phoniatischen Sprechstunde, bei denen ein "reiner" Sigmatismus inter-, addentalis, lateroflexus oder lateralis auffiel, auf eine gleichzeitige Störung der [ʃ]- und [ç]-Bildung, wobei zum Teil Palatogramme angefertigt wurden. Kinder blieben bei unserer Überprüfung unberücksichtigt, um entwicklungsbedingte multiple Lautfehlbildungen sicherheitshalber auszuklammern. Auch seltenere Sigmatismusformen wurden nicht mit in die Untersuchung einbezogen.

Als Ergebnis der Studie kann festgestellt werden, dass die Wahrscheinlichkeit, dass bei einem Sigmatismus addentalis oder lateroflexus auch das [ʃ] oder [ç] falsch gebildet werden, etwa 20 Prozent beträgt. Bei einem Sigmatismus interdentalis beträgt die Wahrscheinlichkeit, dass auch [ʃ] oder [ç] falsch gebildet werden, etwa 15 Prozent.

Wie erwartet ist die Quote der gleichzeitigen Fehlbildungen des [ʃ] und [ç] bei Sigmatismus lateralis recht hoch: 88 Prozent für [ʃ], 84 Prozent für [ç].

Zusammenfassend sollte demnach aus lautphysiologischen und sprachtherapeutischen Gründen der Sammelbegriff "Sigmatismus" mit Zurückhaltung gebraucht werden. Sinnvoller wäre zweifellos eine bereits häufig praktizierte Aufgliederung der Zischlautstörungen in Sigmatismen für die Fehlbildungen des [s] sowie Schetismus bzw. Chitismus für die Störungen der [ʃ]- bzw. [ç]-Bildung.

## L'INTONATION ANGLAISE ET L'ENSEIGNEMENT PAR VISUALISATION

Jennifer Low, Laboratoire de Phonétique

Département de Recherches Linguistiques de l'Université, Paris VII

Une étude instrumentale (spectrographe, mingographe et oscilloscope) de l'intonation anglaise dans le cadre de la théorie de l'énonciation permet des applications dans le domaine pédagogique.

Les analyses instrumentales montrent l'importance de deux phénomènes en intonation ; premièrement la forme de la courbe mélodique à la fin du schéma intonatif, deuxièmement les parties proéminentes à l'intérieur du schéma, les sommets de la courbe.

Les études effectuées montrent que les courbes mélodiques peuvent être reliées à des opérations énonciatives, telle que la modalisation et que les sommets de la courbe sont la trace d'opérations, telle que l'opération de quantification / qualification, ou de relations.

Une étude de l'acquisition de l'intonation anglaise par les étudiants francophones a établi leurs difficultés et leurs fautes et a montré que la symbolique souvent utilisée (points ou tirets) ne correspond pas à la réalité physique et peut même induire l'étudiant en erreur. Il a donc fallu trouver des moyens visuels plus adaptés. Le matériel pédagogique comprend un oscilloscope à mémoire et un magnétophone reliés à un extracteur de mélodie. Le cours permet une rétroaction visuelle aussi bien qu'auditive de la part de l'étudiant.

L'oscilloscope à mémoire transmet la courbe mélodique du schéma intonatif, prononcé par le professeur, sur la partie supérieure de l'écran. L'étudiant répète après le modèle, et la courbe mélodique qui correspond à sa répétition se transcrit sur la partie inférieure de l'écran. L'étudiant compare les deux courbes et fait varier la sienne jusqu'à ce que sa répétition soit correcte.

Nos expériences ont montré que l'auto-correction aussi bien visuelle qu'auditive donnent d'excellents résultats dans l'enseignement des courbes mélodiques anglaises. Néanmoins, les étudiants arrivent difficilement à produire les sommets à l'intérieur du schéma intonatif comme la marque des opérations énonciatives. Ils ne sont pas toujours conscients de ces phénomènes phono-syntaxiques qui sont, pourtant, fondamentaux à la langue anglaise. Il est donc important d'enseigner l'intonation anglaise en liaison avec les opérations énonciatives dont elle est la trace en surface.

NOTES ON THE INTONATION OF SPECIAL AND YES-NO QUESTIONS IN  
ROMANIAN COMPARED TO ENGLISH

Tatiana Makarenko, "Babes-Bolyai" University, Cluj-Napoca,  
Romania

This paper attempts to analyse basic intonation patterns of Romanian special and yes-no questions by means of the auditive method. Comparison being made to English, we have applied the Armstrong-Ward system of notation to Romanian as well.

We have found that of the five elements it is the head (not the nucleus as in English) that is obligatory for any tonogram. In contradistinction to English special questions where the place of the nucleus depends on the speaker's intention, special questions in Romanian do not exhibit any variation, their peculiarity being as follows:

- (a) The stressed syllable of the interrogative word (the head) is always high-pitched;
- (b) All the following syllables (forming the body, nucleus and tail) are uttered on a low-pitched monotone which can be broken neither by the communicative weight of the component words nor by the increasing number of syllables.

The intonation patterns of Romanian and English yes-no questions differ in several points:

- (a) The most important element of the Romanian tonogram - the head - is characterized by an abrupt rise of the voice pitch which remains steady for head and body. It is a static high level pitch. It is opposed to the English nucleus which is notable for a gliding tone-movement from a very low to a higher pitch level (a kinetic tone) if there is no tail. In patterns with tail, the nucleus has the lowest pitch (which is a static tone), while the syllables of the tail gradually ascend;
- (b) The syllables of the body are uttered on a high-pitched monotone in Romanian, while in English they form a gradual descending scale;
- (c) The syllables of the nucleus and tail in Romanian are uttered on a mid-pitched monotone. In English the nucleus and tail can never have the same pitch level.



## THE PERCEPTION OF ENGLISH MINIMAL PAIRS BY GREEK LISTENERS

E. Panagopoulos, Department of English, University of Thessaloniki, Greece

This paper describes a reaction time experiment aiming at scaling eleven English vowels according to perceptual difficulty. The hypothesis is that the degree of variation exhibited in the results reflects degrees of inherent perceptual difficulty experienced by native Greek learners of English during aural discrimination of British English minimal pairs. The longer the relative reaction time, the more difficult the discrimination.

Method

One female native English speaker recorded two sets of carrier sentences. Set A ended in minimal pairs which were selected on articulatory criteria and consisted of adjacent gestures, where the possibility of perceptual confusion might arise. Set B consisted of pairs of identical sentences so that all eleven vowels contained in Set A were tested individually.

Results and Conclusion

The rank order of the means represents degree of perceptual difficulty. 'Same' responses take longer than 'different' responses. There is a close correlation between degree of difficulty and error in judgment in the first top timings. 'Same' sounds are processed in a different manner than minimally contrasted sounds.

References

- Darwin, C.J. (1975): "The perception of speech", Haskins Laboratories Status Report on Speech Research 42/43, 59-102.
- Eimas, P.D. and J.L. Miller (1975): "Auditory memory and the processing of speech", in Developmental Studies of Speech Perception, W.S. Hunter Lab. Psych. Brown Univ. Progress Report No. 3.
- Posner, M.I., M.J. Nissen, and M.R. Klein (1976): "Visual dominance: an information processing account of its origins and significance", Psychol. Rev. 83, 157-189.
- Repp, B.H. (1976): "Posner's paradigm and categorical perception: a negative study", Haskins Lab. Status Report on Speech Research 45/46, 153-167.
- Shankweiler, D., W. Strange and R. Verbrugge (1975): "On accounting for the poor recognition of isolated vowels", Haskins Lab. Status Report on Speech Research 42/43, 277-284.
- Sheldon, D. (1972): "Some implications from a choice reaction time study of speaker discrimination", Univ. of Essex, Lang. Centre, Occasional Papers 13.

## FINAL REPORT ON A STUDY IN GENERATIVE ORTHOGRAPHY

Marc L. Schnitzer, La Universidad de Puerto Rico,  
Rio Piedras, Puerto Rico

The primary contact which many non-native speakers have with the English language is visual. Thus, there exist many competent readers of English who are ignorant of pronunciation. In the past, English pronunciation has been taught in a case-by-case fashion, without regard to principles relating orthography to pronunciation.

This is a report on the efficacy of teaching the pronunciation of English polysyllables to non-native speakers by means of ordered rules which use standard orthographic representations as underlying forms. These rules were tested on two groups consisting mainly of francophones. Both groups were asked to read lists of English words ending in nineteen different suffixes representing fifteen different word classes. The experimental group applied ordered quasiphonological rules to selected words from each of the fifteen word classes being tested. The control group performed repetition exercises on these same words.

A two tailed Mann-Whitney U-test shows that the absolute improvement and relative improvement of the experimental group as compared to the control group are significant at the .02 and .002 level, respectively.

Pedagogical and psycholinguistic implications are noted.

## ENGLISH AND SERBO-CROATIAN VOWEL PHONEMES AND ERRORS MADE

Časlav S. Stojanović, Foreign Language Institute, Belgrade,  
Yugoslavia

This paper presents an analysis of the vowel phonemes of English and Serbo-Croatian, indicating points of agreement and disagreement, coupled with the errors made by most learners of either language at the elementary level courses.

Subject

The English vowel system comprises 7 short and 5 long vowels, as well as 8 diphthongal glides. The Serbo-Croatian vowel system, however, comprises only 5 vowels, which can be long or short, with an additional prosodic feature, that is, a specific pitch change, a choice of four melodic accents.

A few phonemes in the two languages can be said to be phonetically similar. There are great differences in sequence and distribution, and particularly in prosodic features. Whereas in English there are diphthongs and even triphthongs, in Serbo-Croatian there are only vowel clusters. Distributionally, five English vowels cannot occur finally, while all the Serbo-Croatian vowels can occur in all positions. Some of the clusters can only occur non-finally. As for suprasegmentals, the English vowels can differ in length only, while the Serbo-Croatian ones differ both in length and pitch change.

Conclusion

English learners of Serbo-Croatian make errors due to the lack of vowel reduction in the target language, the distributional and suprasegmental features. Serbo-Croatian learners of English are faced with a lot of unfamiliar phones.

References

- Gimson, A.C. (1965): An Introduction to the Pronunciation of English, London: Edward Arnold (Publishers) Ltd.
- Stojanović, Č. (1967): Pronunciation Errors Made by English Learners, Belgrade: Foreign Language Institute.
- Stojanović, Č. (1967): Pronunciation Errors Made by Serbo-Croatian Learners, Belgrade: Foreign Language Institute.

## PERCEPTION AS AN AID IN TEACHING GERMAN PRONUNCIATION

Rudolf Weiss, Department of Foreign Languages, Western Washington University, Bellingham, Washington USA

The speaker has for a number of years concerned himself with a study of various aspects of German vowels, including vowel perception, phonetic and phonemic considerations, and problems in teaching German pronunciation. Results of these efforts have already been reported in previous phonetics and linguistics congresses. These efforts, in particular those in vowel perception, have now also prompted the undertaking of the writing of a new phonetics instruction manual intended primarily for American students of German.<sup>1</sup> This manual differs from other phonetics instruction manuals in its approach since it is based primarily on perception as both a tool and criterion in the presentation and learning of German sounds. This paper thus concerns itself with a description of the principles underlying this approach in applied phonetics in recognizing the role of perception in learning German pronunciation.

The speaker has long asserted that perception differences underlie and parallel production difficulties. This approach rests in large part on this premise. A perception test based on one already developed will aid in establishing particular perceptual difficulties especially for vowels and can be used to measure progress in eliminating certain perceptual errors.<sup>2</sup> Furthermore, results of perception tests given to both native Germans and non-native students of German have influenced greatly the manner in which sounds are treated in this approach, as well as the sequence of their introduction and the nature of the drill materials used. Emphasis is placed as much as possible on the difficulty of each sound from a perception standpoint and on perceptual tendencies and interference difficulties American learners have shown in regard to that sound. Further details of this approach are given in the paper.

- 
- (1) The manual is co-authored by H.H. Wängler and is tentatively titled German Pronunciation: A Phonetics Instruction Manual. The publisher and publication date will be released at the congress.
  - (2) The use of this test in teaching German pronunciation was described in considerable detail in a paper presented by the speaker at IPS-77 (Miami Beach, December 1977) and titled "A Perception Test as a Diagnostic Tool in Teaching German Pronunciation."

## NASALS AND NASALIZATION AS TREATED BY EARLY MUSLIM PHONETICIANS

Muhammed Hasan Bakalla, Phonetics Laboratory, Faculty of Arts, University of Riyadh, Riyadh, Saudi Arabia

This paper attempts to give a summary of the contribution made by early Arabs and Muslims in the field of phonetic sciences. Works by scholars like Al-Khalil (d. 791), Sibawaihi (d. 793), Ibn Jinni (d. 1001), Ibn Sina or Avicenna (d. 1037) and others will be given special attention in this connection. In particular, this paper will present the various treatments of the Arabic nasal sounds and the phenomenon of nasalization.

Nasals as a category of sound

As a term of reference, the Arab and Muslim phoneticians divided the Arabic phonemes into categories such as: glottals, pharyngeals, palatals, dentals /l, r, n/, and labials /f, b, m, w/. Al-Khalil is one of the first Arab phoneticians to order the Arabic phonemes, in terms of place of articulation, along the vocal tract from the glottis upward to the lips. His student, Sibawaihi, and later phoneticians also recognized other categories in terms of manner of articulation such as: voiced/voiceless, stop/non-stop, rolled, lateral, nasals /m, n/. They also recognized nasal variants, e.g. [ŋ, N].

Nasality as a distinctive feature

Further, Sibawaihi and Ibn Jinni seem to lay more emphasis on treating "yunnah" or nasality and other features in terms of binary distinctive feature analysis.

Nasalization as a prosodic feature

The Muslim phoneticians also recognized that in certain contexts /n/ and /m/ may influence non-nasals, both vowels and consonants.

Conclusion

A close look at the early Arabic grammatical works reveals an underlying systematic approach and a rich mine of terminology which are relevant both to modern Arabic phonetics and general phonetics.

References (apart from the original sources)

Semaan, K.I. (1968): "Linguistics in the Middle Ages", Leiden: Brill.

## PRODUCTION AND PERCEPTION OF SPEECH - AN INDIAN ANALYSIS

Moti Lal Gupta, Rajasthan University, MSJ College, Bharatpur, India

This paper attempts to throw light on the Indian way of analyzing speech production and speech perception. According to Indians these two aspects of speech cannot be separated because without the one the other does not exist. The Indians had no access to phonetic instruments and they analyzed on the basis of their own observations which have been noted to be so accurate as to defy many instrumental results.

Subject

Human speech has two well-marked divisions: (1) Production and (2) Perception. According to Indian grammarians and phoneticians communication has a four-fold basis, 'ādhāracatuṣṭaya' - consisting of prayoktā, s'rotā, pratipādyā, and s'abdajñāna - namely the speaker, the listener, the spoken matter, and the knowledge of the spoken matter. According to the Vākyapadīya, the mechanism of speech has to pass through four stages each way to become effective. The process of production begins with 'icchā' or desire to communicate, the next one is the word concept or 's'abdabhāvanā', the third one being effort or 'prayatna', and the final one audible speech or 'uccarāna'. In the case of perception, it begins with the audible sound called 'nāda', which gives birth to 'sphoṭa',<sup>1</sup> and then the words are conceived by 'dhvani' with the result that meaning through 'svarūpa' or form is grasped by the listener.

Conclusion

The Indian analysis has a metaphysical background for they believed in the theory of 'vāka' or speech which is nothing but the manifestation of 'brahma'. The speaker, the listener, the words and the meaning are all emanations from the ultimate word-principle, 's'abdatattva'.

References

- Allen, W.S. (1953): Phonetics in Ancient India, Oxford.  
 Iyer, K.A.S. (1963): Vākyapadīya of Bhartṛhari with the Vṛiti, Poona.  
 Kṛṣṇamācārya, V. (1947): Nāgojī's Sphoṭavāda, Madras.  
 Varma, S. (1929): Critical Studies in the Phonetic Observations of Indian grammarians, London.  
 Whatmough, J. (1956): Language, London.

(1) Sphoṭa has been translated in several ways - breaking forth, splitting open, bursting, disclosure, etc. It is the impression produced on the mind and the form that is created before us.

## GRAPHEMICS AND THE HISTORY OF PHONOLOGY

J.H. Hospers, Institute of Semitistics, State University of Groningen, Netherlands

It has often been said that the invention of the alphabet marked the beginning of phonology, of course meant in a pre-scientific sense of the word. On the other hand, it has also been said that "the phoneme concept would never have been developed without the alphabetic script" (F. Balk-Smit Duyzentkunst, 1978, p. 2). Apart, however, from this question the assumption of Mrs. Balk is in itself a sign that something has altered in linguistics during the last decennia. Writing has come into the picture in linguistics again. Linguistic structuralism was based exclusively on spoken language, but thanks to the works of such scholars as H.J. Uldall, J. Vachek and W. Haas the insight has grown that writing is worth studying also linguistically.

Now, however, the following question arises: What remains of the above mentioned relation between phoneme and alphabet (or vice versa), if T.G.G. is right in saying that a phonological level corresponding to a psycho-linguistic reality does not exist (cf. Chomsky and Halle 1968)? Mrs. Balk has proved in her article that, in any case, the notion phoneme does exist and that certain specifications and elaborations of this concept contain alphabetic elements. T.G.G. acknowledges only a morphonological and a phonetic level, both levels being connected with each other by a set of general rules. So most orthographies are morphonological and not merely phonological. The history of writing, now, also teaches us that not only the alphabetic but also the syllabographic scripts were from the beginning already of a morphonological kind. Especially such grammatologists and semitists as I.J. Gelb and E. Reiner have drawn attention to the use of what they call: "morphographemics" (= morphophonemic spellings) in the ancient writing systems. So, perhaps, we have to conclude to some unconscious activity of a morphophoneme concept in the human mind.

#### References

- Balk-Smit Duyzentkunst, F. (1978): "Phoneme and alphabet", in Linguistics in the Netherlands 1974-1976, 1-6, W. Zonneveld (ed.), Lisse.
- Chomsky, N.A. and M. Halle (1968): The Sound Pattern of English, New York: Harper and Row.

## 19TH-CENTURY ATTEMPTS AT THE CREATION OF A UNIVERSAL PHONETIC ALPHABET: FROM VOLNEY TO PASSY

E. F. K. Koerner, Dept of Linguistics, Univ of Ottawa, Ottawa, Canada K1N 6N5

This paper surveys the various attempts to establish a universal phonetic alphabet during the period between Volney's *Alphabet européen* (1819) and the foundation of the International Phonetic Association by Paul Passy, Henry Sweet, Otto Jespersen, and others in 1888 and the subsequent creation of an 'international' phonetic alphabet.

Emphasis is placed on those 19th-century approaches that have been either totally ignored or treated inadequately in Albright's (1958) account of the pre-history and background of the IPA. Albright, for example, is heavily bi-ased in favour of the British tradition(s); not only is his chapter entitled "Early Backgrounds" exclusively devoted to 16th and 17th century phoneticians in England (John Hart, Wilkins, Holder, and others), but also the subsequent chapter entitled "Nineteenth Century Backgrounds" deals almost exclusively with Anglo-Saxon efforts in the field, in particular the work of A. M. Bell, A. J. Ellis, I. Pitman, and Henry Sweet.

It appears that the first scholar to take up Volney's suggestions was A. A. E. Schleiermacher (1787-1858), who in 1835 published a 700-page study on writing systems, to which he added an "Alphabet harmonique pour transcrire les langues asiatiques en lettres européennes". (Pickering's proposals of 1818 were made independently of Volney's, but they suggest that the development of a universal phonetic alphabet was 'in the air' at the beginning of the 19th century.) In 1864 Schleiermacher published a revised German version of his earlier proposal; by that time, however, many other attempts had been made throughout Europe to develop universal systems of phonetic transcription of non-Indo-European languages, of which the following authors may be taken as representative (ignoring the British contribution to the field): Alexandre Erdan's (alias Alexandre André Jacob, 1826-78) *Congrès linguistique: Les révolutionnaires de l'A-B-C* (Paris, 1854); Richard Lepsius' (1810-84) *Das allgemeine linguistische Alphabet* (Berlin, 1855; 2nd rev. E. ed., London, 1863); Felix Heinrich Du Bois-Reymond's (1782-1865) *Kadmus, oder allgemeine Alphabetik* (Berlin, 1862), and perhaps also Paul Jozon's (1836-81) *Des Principes de l'écriture phonétique et des moyens d'arriver à une écriture universelle* (Paris, 1877).

The present paper attempts to redress the balance of previous scholarship and hopes to make a contribution to the historiography of phonetics,

#### References

- Albright, R. W. (1958): The International Phonetic Alphabet: Its backgrounds and development. Bloomington: Res. Ctr. in Anthropol., Folklore & Linguistics.  
 Austerlitz, Robert (1975): "Historiography of Phonetics: A bibliography". In: Current Trends in Linguistics, vol.13.1179-1209. The Hague: Mouton.



UN ESSAI ORIGINAL DE TRANSCRIPTION MUSICALE DE LA PROSODIE:  
WILHELM WUNDT: VÖLKERPSYCHOLOGIE (1900 - 1912)

Gabrielle Konopczynski, Laboratoire de Phonétique, Université de Besançon, France

L'étude porte sur la tentative de notation musicale de la prosodie proposée par Wundt dans sa Völkerpsychologie, car, à notre connaissance, les ouvrages consacrés à l'historique de cette question ignorent totalement cet auteur, dont la démarche nous paraît originale à quatre titres:

1) par la place du chapitre sur la prosodie, inséré dans la section sur la syntaxe, fait intéressant à signaler à notre époque où les linguistes essaient de tenir compte de la prosodie pour l'élaboration de leurs théories syntaxiques.

2) par l'optique particulière dans laquelle se place le créateur de la psychologie expérimentale. Tout en reprenant à son compte les travaux de ses contemporains phonéticiens, Wundt étudie surtout le décalage entre l'aspect physique des phénomènes et leur aspect perceptuel subjectif.

3) par la priorité donnée à la perception qui l'amène à traduire en termes musicaux, à l'aide d'une notation très personnelle, les variations prosodiques: aux notes habituelles sont ajoutées des lignes mélodiques soulignant le caractère continu des changements de hauteur (procédé inspiré des "Intonation Curves" de Jones?); sont également visualisées intensité et durée, grâce à une graduation horizontale de la portée en intervalles de temps égaux.

4) par la réflexion linguistique de Wundt qui dégage la fonction linguistique oppositive des éléments prosodiques, en montrant comment une même phrase peut prendre diverses modalités (énonciatives, ordre, question, condition ...) selon l'agencement des divers paramètres prosodiques.

Appuyée sur une parfaite connaissance des phénomènes physiques et perceptuels, la réflexion linguistique de Wundt, avec sa résonance fort moderne, constitue une contribution que l'histoire de la phonétique ne saurait ignorer.

Références:

- Roudet, L. (1910): Eléments de phonétique générale. Paris: Welter.  
Rousselot, Abbé P.J. (1901-1908): Principes de phonétique expérimentale. Paris: Welter.  
Sievers, E. (1876): Grundzüge der Phonetik. Leipzig: Breitkopf.

## REGARDS SUR L'HISTOIRE DE LA PHONÉTIQUE ET DE LA PHONOLOGIE

András O. Vértes, Institut de Linguistique de l'Académie Hongroise des Sciences, Budapest

1. L'histoire de la phonétique et de la phonologie est loin d'être élaborée. L'histoire de notre discipline ignore, par exemple, quel était le rôle articulatoire attribué aux dents et à la langue dès l'Antiquité jusqu'à la fin du Moyen Âge. Nous prouvons à l'aide de nombreux auteurs -- de Cicéron jusqu'à Barthélemy l'Anglais -- qu'il existait une conception étrange au sujet de l'articulation.

2. Les deux grands domaines de la science de la voix articulée, notamment ceux de la phonétique et de la phonologie étaient déjà distingués par Simon Dacus, au XIII<sup>e</sup> siècle.

3. Le concept phonologique du phonème existait au cours des siècles passés, sous une forme plus ou moins consciente.

4. Certains moments de l'histoire de la phonétique illustrent la manière de penser des époques respectives. Au Moyen Âge, on supposait une liaison réelle entre le nom (la chaîne phonique) et le concept qu'il désigne, on croyait à une connexion réelle entre le symbole et le concept symbolisé, ainsi entre la lettre et le son.

5. C'est le moment historique favorable qui explique l'épanouissement de la phonologie dans notre époque; c'est également le moment favorable qui rend compréhensible l'influence exercée par la phonologie sur d'autres domaines de la linguistique et des sciences humaines.

## FRÜHNEUENGLISCHE WEGE DER LAUTBESCHREIBUNG

Horst Weinstock, Technische Hochschule, D-5100 Aachen

Der frühneuenglischen Lautbeschreibung fehlt es an Eindeutigkeit. Nach dem Vorbild farbiger Auszeichnung bei dunkler Grundschrift in mittelalterlichen Manuskripten erfüllten im elisabethanischen Buchdruck Fraktur, Antiqua und Kursive zunächst ästhetische Funktion. Die Buchart (klassisch, liturgisch, rechtlich, scholastisch, volkstümlich) regelte die Wahl der Grundschrift. Bei mehreren Buchteilen wechselte der Satz und untergliederte in Widmung, Vorwort, Einleitung, Hauptteil und Nachwort. Von Fall zu Fall griff der elisabethanische Setzer zu Auszeichnungsschrift als Lesehilfe ohne sprachwissenschaftliche Aufgaben.

Die vorrangige Pflege klassischer Schriftsprache(n) vertiefte die Denkabhängigkeit vom lateinischen Alphabet. Phonographisch entsprach sein Inventar kaum der frühneuenglischen Lautung.

Der Vortrag behandelt die fünf frühneuenglischen Verfahren der Lautbeschreibung: (1) orthographische Vereinheitlichung, (2) komparatistische Umschreibung, (3) numerische Verfeinerung der Buchstaben, (4) allotypische Auszeichnung und (5) diakritische Anpassung des Alphabets. Orthographische, komparatistische und numerische Lautbeschreibungen schieden als ungenau und umständlich aus. Allotypische und diakritische Lautwiedergaben leben in enger Umschrift fort. Als weite Umschrift bewährte sich das type-token-System. Setzung oder Nichtsetzung von eckigen Klammern oder Schrägstrichen signalisiert die Typen phonetische ~ phonemische Lautung bzw. Schreibung. Für die Lautwerte stehen alphabetische Zeichen.

Die frühneuenglische Transkription schritt in drei Stufen voran: Die Schreibreformer des 16. Jahrhunderts versuchten 'geschriebene' Buchstaben lautgetreu zu regeln. Die Phonetiker des 17. Jahrhunderts widmeten sich der Aussprache der Buchstaben; sie unterschieden dabei nicht scharf genug zwischen Buchstabennamen des Alphabets und 'gesprochenen' Buchstaben im Redefluß. Erst den Orthoepisten des 18. Jahrhunderts gelang der gedankliche Durchbruch vom Buchstaben zum Laut. Bei aller Unvollkommenheit und Unausgeglichenheit entwickelten sie echte Transkription, das heißt 'geschriebene' Laute.

Quellen

English Linguistics 1500-1800 (A Collection of Facsimile Reprints),  
Selected and Edited by R.C. Alston, Leeds/Menston, 1967ff.

## PHONOLOGY IN SOCIOLINGUISTICS: WHAT DO THE DATA TELL US?

Lawrence M. Davis, University of Haifa, Haifa, Israel

The purpose of this paper is to raise certain questions about the validity of some by now well established principles for the analysis of the phonological data in sociolinguistic studies. The data for the paper come from a study of the language of disadvantaged Israeli schoolchildren and, because of limitations of time and space, the records of sixteen respondents are analysed.

The paper concludes that the presentation of data in the form of graphs which show the mean and/or median percentages of the incidence of phonological variables may indeed be misleading. When the mean results of our study are graphed in the usual way, we seem to get clear class stratification of our variables, but when standard deviations are calculated the results are far less clear cut. Similarly, phonological variation as a function of register, or contextual style, may be graphed quite neatly. Yet standard deviation again makes the findings far more difficult to analyse.

It is argued here that sociolinguistic studies must present more than the mean and/or median percentages; they must also include data on standard deviation as well. The results might not be as neat that way, but our analyses should tell us more of what we want to know about language.

INFLUENCE DU FACTEUR SEXUEL SUR L'ARABE MAROCAIN D'OUDJA  
(MAROC ORIENTAL) - (ILLUSTRATION PHONOLOGIQUE, SYNTAXIQUE,  
LEXICALE)Simone Elbaz, Paris

Pour analyser l'influence du paramètre différence de sexe au plan de la phonologie - et aussi de la syntaxe et du lexique - nous avons rapproché les productions de deux sujets parlants oujdis de sexe différent mais comparables du point de vue d'autres facteurs externes tels l'origine des parents, le lieu de naissance, l'âge, la première langue acquise, la stabilité géographique, le milieu socio-professionnel. Les productions de ces deux témoins - dits de référence - ont été vérifiées auprès d'autres témoins présentant les mêmes caractéristiques durant nos différentes missions sur le terrain. Ainsi pouvons-nous nous autoriser à qualifier cette variable dans l'optique d'une synchronie dynamique et prendre position quant au caractère conservateur ou novateur du comportement linguistique des femmes oujdies.

En effet l'analyse des données recueillies révèlent que les différences de réalisations entre locuteurs et locutrices ne sont pas toutes caractéristiques de l'un ou de l'autre sexe. Les deux ont à leur disposition les mêmes traits distinctifs, les mêmes phonèmes. Pour les autres phénomènes envisagés nous parlerons de tendances que l'accès de plus en plus important des femmes oujdies à la vie sociale semble devoir unifier.

## A SOCIOLINGUISTIC APPROACH TO THE PROBLEM OF NORMALIZATION

William Labov, University of Pennsylvania, Philadelphia, Pa., USA

The measurement of sound change in progress can be advanced considerably by recent techniques for formant analysis such as LPC. But increased accuracy will not help in placing trends across age level unless progress is made towards solving the normalization problem, so that changes in mean vowel position can be related to a single reference grid.

The normalization method that shows the greatest clustering is not necessarily the best, since significant characteristics of the data such as age-grading can be removed by too powerful clustering techniques. Optimum normalization will eliminate only those acoustic differences due to differences in vocal tract length. The preservation of social differentiation that is independent of vocal tract length offers the most decisive test of a normalization method.

Measurements of vowel systems of 176 Philadelphians were submitted to three normalizations: the vocal tract model of Nordström and Lindblom (1975); the log mean model of Nearey (1977); and the six parameter regression of Sankoff, Shorrock and McKay (1974). It can be shown that the Sankoff model is too powerful, and that both the Nearey and Nordström & Lindblom normalizations preserve socio-linguistic relations that are masked in the unnormalized data and eliminated by the very high degree of clustering achieved in the Sankoff normalization.

#### References

- Nearey, T. (1977): Phonetic feature systems for vowels. Unpublished University of Connecticut dissertation.
- Nordström, P.-E. and B. Lindblom (1975): "A normalization procedure for vowel formant data", Paper 212 at the 8th Int. Cong. of Phonetic Sciences, Leeds.
- Sankoff, D., R.W. Shorrock and W. McKay (1974): "Normalization of formant space through the least squares affine transformation", Unpublished program and documentation.

## BB. CONTRIBUTION A L'ETUDE DES VOIX DE CHARME

Pierre R. Léon, Laboratoire de phonétique expérimentale,  
Université de Toronto

On propose ici une analyse de la voix de Brigitte Bardot dans une situation de communication bien précise, une interview pour grand public. Un groupe d'auditeurs a attribué à cette voix le qualificatif de voix de charme avec les sous-catégories: amoureuse, coquette, petite fille. On retrouve pour les deux premières sous-catégories, un certain nombre de traits déjà décrits par Moses, Trojan, Fónagy et Magdics, Fónagy, etc. Le charme amoureux se manifeste par les traits vocaux: atténuation d'intensité et souffle; et par les traits prosodiques: ralentissement et décélération du tempo, patron mélodique descendant. Le charme de la coquetterie se signale par une montée mélodique abrupte sur la finale de groupe rythmique. La voix du charme petite fille ne semblait pas avoir été décrite de manière précise. Elle se manifeste chez BB par des traits articulatoires: antériorisation et fermeture; et par des traits prosodiques: accélération du tempo, préférence pour les groupes rythmiques courts et surtout par un changement total du patron rythmique au niveau syllabique par rapport au français standardisé.

Il semble que le degré de conscience des processus métaphoriques employés varie en fonction du nombre de traits nécessaires à la description.

## ANOTHER LOOK AT STAGES IN THE ACQUISITION OF STANDARD ENGLISH

Renate Portz, Institut für Englische Philologie, Freie Universität Berlin, 1000 Berlin 33 (West Germany)

The age-range between child- and adulthood has been relatively neglected in socio- and psycholinguistics. This paper presents an attempt to examine in more detail the developmental phases in phonological variation and attitudes towards language varieties of 9 to 18 year olds. The prevailing concept of a continuous linguistic and metalinguistic acculturation in terms of a gradually increasing conformity to adult norms is challenged on the basis of the findings of an empirical study of youths in Norwich, England. Patterns of phonological variation and evaluation of both male and female speakers of Standard and Nonstandard English (elicited in a "matched-guise test") show that there is an interval of significant regression in the stages of acquisition of Standard English. The group of 15 to 16 year olds strongly reject the Standard norm by both their actual linguistic behaviour as well as their partly unconventional attitudes towards nonprestigious speech varieties. These findings are tentatively discussed in the frame of interactional developmental psychology as linguistic and metalinguistic correlates of sex-role identification and identity formation processes in adolescence.



## L'ORGANISATION POLYLECTALE DE LA PHONOLOGIE

Gilbert Puech, UER Sciences du Langage, Université Lyon II, France

Cet exposé montre comment organiser une phonologie polylectale dont l'objectif soit de décrire le système propre à chaque locuteur dans sa relation avec le dia-système de la langue qu'il parle. Une telle organisation permet notamment une définition linguistique de ce qui constitue un idiolecte, un dialecte, une langue.

La méthode (appliquée ici au maltais) consiste à considérer chaque ensemble fonctionnel de relations paradigmatiques ou syntagmatiques comme une lecte autonome et de décrire l'organisation de chaque lecte comme un micro-système de relations et de règles du type de celles utilisées en phonologie générative.

Quatre lectes du maltais sont présentées pour illustration:

Lecte A: Voyelles brèves (organisations de l'espace vocalique)

Lecte B: L'harmonie vocalique

Lecte C: La syncope d'une voyelle brève et l'épenthèse d'un i

Lecte D: L'abrègement des voyelles longues et l'accent.

L'étude de cet exemple montrera comment un ensemble de lectes circonscrit l'espace phonologique d'une langue et comment la combinaison des options à effectuer pour chaque lecte permet d'analyser un ensemble ouvert d'idiolectes. On verra également que ces choix se lisent "en diachronie" comme des changements (en cours ou déjà effectués) et en "synchronie" comme des variantes (avec ou sans valeur socio-linguistique).

#### Conclusion

Les variations dans une langue sont dues, de façon indirecte et médiatisée, à l'hétérogénéité de la communauté qui utilise cette langue. La théorie linguistique doit prendre en charge le réseau d'interférences qui en résulte pour être explicative.

#### Bibliographie

Bailey, C.J. (1973): Variation and Linguistic Theory, Arlington, Virginia: Center for Applied Linguistics.

Chomsky, N. et M. Halle (1968): The Sound Pattern of English, New York: Harper and Row.

Haudricourt, A. et C. Hagège (1978): Traité de phonologie pan-chronique, Paris: P.U.F.

Labov, W., M. Yaeger et R. Steiner (1972): A quantitative study of sound change in progress, Philadelphia: The U.S. Regional Survey.

Puech, G. (en préparation): "Les parlers maltais; essai de phonologie polylectale".

## THE NOTION OF INTERMEDIATE PHONETIC FORMS IN SOCIOLINGUISTICS

Suzanne Romaine, Department of Linguistics, University of Edinburgh, Edinburgh, Scotland

The notion of so-called "intermediate forms" is not confined to sociolinguistics, but it is implicit in Labov's (1966) definition of the linguistic variable, which is the starting point for many sociolinguistic analyses. Labov (1966, 15) has defined the linguistic variable as a class of variants which are ordered along a continuous dimension and whose position along that dimension is determined by some independent or extralinguistic variable. This concept assumes, among other things, a continuum, and hence, intermediate stages between one end of the continuum and the other.

Although a number of phonetic and phonological variables are handled quite easily within such a gradient framework, others are not, and are better considered as discontinuous or discrete. In cases where variation may be treated either as discontinuous or continuous, the construction of variable scales is often done without consideration of the extralinguistic nature of the variation, i.e. how heterogeneous it is in a given speech community.

Using evidence from variation in word final /r/ in Scottish English, I will attempt to show that the fact that variation can be observed and described by a continuous process with intermediate stages, does not demonstrate that the variation cannot also be generated by an underlying discontinuous model. Furthermore, I will argue that quantitative evidence is not always a sufficient basis for deciding whether or not a linguistic process is continuous or discontinuous or whether one rule or separate rules is/are involved.

Reference

Labov, W. (1966): The Social Stratification of English in New York City, Washington D.C.: Center for Applied Linguistics.

REGRESSION RAPIDE DU [r] APICAL DANS LE FRANÇAIS DE MONTREAL;  
ETUDE SOCIO-PHONETIQUE

Laurent Santerre, département de linguistique, Université de Montréal

A partir de 51 interviews d'une heure de locuteurs franco-phones de Montréal (corpus Sankoff-Cedergren), on a étudié la distribution et la variation des /R/ selon les groupes d'âge, le sexe, les classes sociales et le rang sur le marché linguistique (Bordieu 1977, Laberge 1978). L'analyse et les corrélations des 15294 /R/ relevés permettent de constater un changement rapide en faveur des variphones postérieurs; les facteurs les plus déterminants sont dans l'ordre: l'âge, le sexe et le niveau de langage. A l'intérieur de chaque idiolecte, les consonnes en coarticulation et la nature des frontières syllabiques semblent avoir une influence contraignante sur le choix du variphone, et ces contraintes n'ont pas le même poids pour chaque locuteur. Certains locuteurs font à la fois une constriction postérieure et une occlusion apicale pour la réalisation d'un même R dans un mot. L'étude articulatoire permet de proposer une hypothèse nouvelle (différente de celles de Delattre (1966) et Martinet (1962)) sur la postériorisation du /R/. En 1950, Vinay considérait les R postérieurs comme exceptionnels dans la région de Montréal; en 1971, ils sont passés de presque zéro à 51%. On peut entrevoir les causes de ce changement rapide.

Références

- Bourdieu, P. (1977): "L'économie des échanges linguistiques", Rev. Langue française, 34, 17-34, Paris: Larousse.
- Delattre, P. (1966): "A contribution to the history of "R grasseyé", in Modern Language Notes, 562-564, 1944, reprinted in Studies in French and Comparative Phonetics, 206-208, The Hague:Mouton.
- Laberge, S. (1978): Etude de la variation des pronoms sujets définis et indéfinis dans le français parlé à Montréal, thèse de Ph.D., dép. d'anthropologie, Université de Montréal.
- Martinet, A. (1962): "R, du latin au français d'aujourd'hui", Phonetica 8, 193-202, et Le Français sans fard, 132-143, Paris: P.U.F. (1969).
- Sankoff, D., G. Sankoff, S. Laberge et M. Topham (1976): "Méthodes d'échantonnage et utilisation de l'ordinateur dans l'étude de la variation grammaticale", in La sociolinguistique au Québec, 85-126, Les Presses de l'Université du Québec.
- Santerre, L. (1976): "Nombreux variphones du /R/ en français du Québec", Actes du troisième congrès mondial de phonétique, Tokyo, août 1976.
- Vinay, J-P. (1950): "Bout de la langue ou fond de la gorge?", in The French Review 23, 489-498.

## DIGITALE PHONETISCHE ANALYSEN FÜR "VERHALTENSPARTITUREN"

Peter Winkler, Forschungsprojekt "Kommunikation und Interaktion" (Leitung: Prof. Dr. Th. Luckmann/Dr. P. Gross) am Fachbereich Psychologie/Soziologie der Universität Konstanz, Bundesrepublik Deutschland

Für die phänomenologische Analyse von unmittelbaren dyadischen Interaktionen ist es sinnvoll, sich einer beschreibenden "Partitur" zu bedienen, die möglichst detailliert und komplex ist. Um dieses Ziel zu erreichen, bedarf es einer multidisziplinären Erforschung von Sprache, Sprechen, Mimik und Gestik. Die Aufgabe des Phonetikers in diesem Team ist es, aus dem Methodenrepertoire der modernen Phonetik solche Verfahren auszuwählen und anzuwenden, die eine exakte Messung, eine umfassende Notierung und eine plausible Visualisierung erlauben, und die sich mit nicht-phonetischen Analysen harmonisch verknüpfen lassen. Folgende Methoden erfüllen diese Anforderungen: normalphonetische Transkriptionen, Notationen der paralinguistischen Merkmale und digitale Messungen bzw. Kurvenaufzeichnungen. Um eine fortlaufende Messung und Registrierung mit mehreren Parametern über eine Gesamtlänge von 120 Min. Sprachschall ökonomisch durchführen zu können, wurde ein digitales Schallanalyseprogramm verwendet.<sup>1</sup> Folgende Grundparameter wurden ausgewählt: Zeitachse in 1/10 Sek., Bildzählung in 1/25 Sek., Oszillogramm, Signal-Pausen-Ratio, Weglänge des Signals, Lautstärkepegel, Grundtonbewegung, Frikativ- und Stimmhaftigkeitserkennung sowie ein digitales Sonagramm. Das digitalisierte Schallmaterial wird zunächst segmentiert und maschinell mit Transkriptionssymbolen versehen (mittels gesondertem Transkriptions-File), damit der Plotter-Ausdruck zugleich Messkurven und Transkription enthält. Im Referat werden die Entwicklung und Anwendung des Analyseprogramms erläutert, der Einsatz phonetischer Methoden in den Sozialwissenschaften diskutiert und einige Beispiele aus dem Analysematerial demonstriert. Ausführlich besprochen wird das entscheidende Problem der Verknüpfung von Transkriptionen und Eindrucksurteilen mit den phonetischen Messungen, nonverbalen Kodierungen und den Daten der Konversationsanalyse.

---

(1) Die Analyseprogramme wurden entwickelt in Zusammenarbeit mit Professor Dr. H.G. Tillmann, Institut für Phonetik und sprachliche Kommunikation der Universität München.

## THE STUDY OF MANSI (VOGUL) VOWELS BY MEANS OF X-RAY PHOTOGRAPHY

Yuri A. Tambovtsev, Novosibirsk State University, USSR

This paper deals with the results of the X-ray photography investigation of the long vowels in the Northern dialect of the Mansi (Vogul) language.

Subjects

Both dynamic and static X-ray photography were used to study the articulation of 5 native speakers - representatives of 4 different subdialects of the Northern dialect of Mansi. The long vowels were taken in one and two syllable Mansi words.

All the measurements of the articulators' shapes and positions are given not in absolute but in relative numbers. It is proposed here to give the measurements in relation to "L<sub>const</sub>", which is the distance from the end of the hard palate to the edge of the front upper teeth, and to divide the tongue into 5 parts.

The analysis of the X-ray photographs of the Mansi long vowels (Srednesosvinski goror) allows to state that they have the following articulatory characteristics (zones, height, labialization):

Russian letters	vowel sound	IPA	articulatory characteristics
a	« ä: $\frac{V}{\text{I}}$ »	[ä]	back zone, very much advanced, low (the V-th grade), non-labialized.
o	« ō: $\frac{V}{\text{I}}$ »	[o+]	back zone, very much advanced, low (the V-th grade), labialized.
y	« ŷ: $\frac{II}{\text{I}}$ »	[ü+]	back zone, very much advanced, high (the II-d grade), labialized.
e, э	« ε: $\frac{IV}{\text{I}}$ »	[ë]	front zone, very much retracted, mid (the IV-th grade - a little closed), non-labialized.
и	« I: $\frac{II}{\text{I}}$ »	[i+]	front zone, a little retracted, high (the II-d grade), non-labialized.

Conclusion

The results of the X-ray photography of Mansi (Vogul) long vowels allow us to obtain the most objective data of their articulation zones and tongue height. It gives a good basis for comparing the articulatory characteristics of the Mansi long vowels to the analogical vowels in other languages, and especially languages of the Finno-Ugric family.

THE PERCEPTION OF CHINESE SPEECH SOUNDS IN MASKING NOISE  
AND FREQUENCY DISTORTION

Tze-Wei Pao and Yung-Tzue Wei, Acoustical Institute of Nanking  
University

Intelligibility tests of Chinese speech sounds were run under five masking conditions, namely white noise, pink noise, speech noise, meaningful speech interference, and reverberation masking in an auditorium, as well as in a quiet studio. To simulate the actual communication circumstances, the noise was introduced at input and output ends, respectively. The signal to noise ratios were 5, 0, -5, -10 dB with a fixed speech level about 80 dB at 1m from the loudspeaker. In addition, the speech and noise were processed with high pass, low pass, or band pass filtering except in the reverberation condition. A set of simplified but rather sensitive word lists were used, which were based on varying the initial consonants (initial consonants are more sensitive to masking than are final consonants). The effects of masking and frequency distortion on the perception of individual Chinese speech sounds will be presented in this report.

PHONETIC UNIVERSALS IN PHONOLOGICAL SYSTEMS AND THEIR EXPLANATION

Summary of Moderator's Introduction

John J. Ohala, Phonology Laboratory, Department of Linguistics,  
University of California, Berkeley, California, U.S.A.

In many ways the study of phonetic and phonological universals is a relatively old endeavor in linguistics and in other ways it is relatively new. It is old in the sense that some 100 years ago when our intellectual forefathers, Ellis, Sweet, Bell, Lepsius, Passy, Jespersen, and others were struggling to develop a workable phonetic alphabet that could be used to transcribe the sounds of any language, they had, implicitly at least, to deal with the problem of whether there were phonetic universals. That they succeeded in devising such a practical universally-applicable phonetic notation, such as we use today, is a tribute to their hard work, their vast experience with languages of the world, and their scientific judgment. The phonetic alphabet and the set of descriptive terms accompanying it are not perfect, of course, but modern work on universal sound patterns would be impossible without these important tools.

It is an old interest, too, in the sense that during the past century there has been a steady, if small, flow of relatively sophisticated explanations for observed universal sound patterns, e.g., (to mention just a few) Passy 1890, Issatchenko 1937, Troubetzkoy 1939, Jakobson 1941, Hockett 1955, Martinet 1955. Characteristic of the keen insights offered during this period are the following (and here I present remarks relevant to topics of particular interest to this symposium):

Passy (1890) on obstruent devoicing: 'On remarque que ce sont les explosives qui se dévocalisent le plus souvent. Cela se conçoit, car pour produire une explosive vocalique, il faut chasser dans une chambre fermée l'air qui fait vibrer les cordes vocales; action qui nécessite un effort considérable, et ne peut pas se prolonger. Aussi les explosives doubles sont-elles particulièrement sujettes à devenir soufflées..' [161].

Chao (1936) on the patterning of voiced glottalized stops: 'A very significant circumstance about the occurrence of [ʔb, ʔd, ʔs, ʔj] is that in all the [Chinese] dialects in which they are known to exist they are always limited to labials and dentals and never exist in velars ... The reason is not

far to seek. Between the velum and the glottis, there is not much room to do any of the tricks that can be done with the larger cavity for a b or a d. As soon as there is any vibration of the vocal cords, the cavity for a g is filled and a positive pressure is created. There is therefore no space or time to make any impression of suspension [of voicing via simultaneous glottal closure, as with ?b and ?d] or of inward "explosion" as with [b, d]. The velar plosive is difficult to voice without having to do any additional tricks.'

In another sense, however, interest in universal sound patterns is rather recent or at least renewed. This has come about, I think, due to the interaction of a number of trends and events. First, there is the interest in phonological universals stimulated by the new set of research goals presented by generative phonology, namely, to look at universals of language for what they will reveal about humans' genetically-based capacity for language.

Second, there has been the sheer accumulation of reliable phonological data on a large number of languages. Works such as Guthrie's Comparative Bantu and Li's Comparative Tai (to mention just two), which synthesize large amounts of phonological data, exemplify this trend. It is because of this latter development that a project such as the Stanford Phonology Archive, constructed by Charles Ferguson and Joseph Greenberg, was possible. Third, there has been the almost explosive growth of experimental phonetics over the past 30 years or so -- especially in the development of empirically-validated mathematical models of various aspects of the speech production and perception mechanisms. In short, phonologists have realized that the study of universal sound patterns can be interesting and very important and that they now have the resources to do a better job of it than ever before.

The contributions to this symposium on phonetic universals represent very well the wide range of data, of talents, and of theoretical outlooks that are necessary in this area.

Björn Lindblom, in 'Some phonetic null hypotheses for a biological theory of language' raises the possibility that the form of language and the range over which it varies when it changes, may be determined by the biological constraints of its human users. He looks to phonetics to provide the evidence on this issue.

Kenneth N. Stevens, in 'Bases for phonetic universals in the properties of the speech production and perception systems' considers how the natural classes among speech sounds must arise due to the individual members of the classes sharing common modes of production at the neuromuscular level and/or giving rise to a common set of sensory images via the tactile, kinesthetic, or auditory channels.

Kenneth Pike, in 'Universals and phonetic hierarchy' suggests that the inability of phonologists to integrate such elusive units as the syllable or stress group into their descriptions of language may be due to their commitment to use just a single hierarchical structure. He proposes the use of parallel but interlocking hierarchies, e.g., one each for the phonological, grammatical, and referential domains.

Two papers in this symposium and one section paper deal with closely related topics on universal patterns in languages' obstruent inventories.

Thomas V. Gamkrelidze, in 'Hierarchical relations among phonetic units as phonological universals', presents a comprehensive analysis of universal co-occurrence tendencies among various features of obstruents, e.g., place of articulation, voicing, glottalization, and uses this to support a reanalysis of the Indo-European stop inventory.

André G. Haudricourt, in 'Apparition et disparition des occlusives sonores préglottalisées', presents the phonetic factors that lead to the development or loss of voiced preglottalized stops and presents extensive supporting cross-linguistic data, especially from South and Southeast Asian languages.

Sandra Pinkerton, in her section paper 'Quichean (Mayan) glottalized and non-glottalized stops: a phonetic study with implications for phonological universals', presents instrumental data on the manner of production of glottalized stops in five Mayan languages. Having found voiceless uvular implosives, she proposes a revision of Greenberg's (1970) implicational hierarchy for glottalized stops which would equate it to Gamkrelidze's claims: voicing is marked for velar obstruents, voicelessness for labial obstruents.

Robert K. Herbert, in 'Typological universals, aspiration, and post-nasal stops', points out several universal patterns characteristic of nasal + stop clusters and uses these to call into



question one reconstruction of the history of such clusters in Eastern Bantu languages.

Jean Marie Hombert, in 'Universals of vowel systems: the case of centralized vowels', presents data from speech perception tests conducted in the field with speakers of Fe?fe? (a Bantu language of Cameroon) which suggest that the universal tendency of disfavoring central vowels may have its origin in a human auditory constraint.

In my own paper, 'Universals of labial velars and de Saussure's chess analogy', I present four phonetically-based universal patterns characteristic of labial velars and use this to call into question the wisdom of structuralist phonology's pre-occupation with system-internal relations in language and their descriptions.

#### Conclusion

It is worth mentioning that study of phonological universals is of more than theoretical interest. If it is done well, it could yield results of great practical benefit, too, e.g., in such areas as speech therapy, second language teaching, speech recognition, and neurophysiology.

#### References

- Chao, Y.R. (1936): "Types of plosives in Chinese", Proc.Phon. 2, 106-110, D. Jones and D.B. Fry (eds.), Cambridge: The Univ. Press.
- Greenberg, J.H. (1970): "Some generalizations concerning glottalic consonants, especially implosives", IJAL 36, 123-145.
- Hockett, C.F. (1955): A manual of phonology, IJAL Memoir 11.
- Issatschenko, A. (1937): "A propos des voyelles nasales", Bull. Soc. Ling., Paris 1938, 267ff.
- Jakobson, R. (1941): Kindersprache, Aphasie und allgemeine Lautgesetze, Uppsala.
- Martinet, A. (1955): Economie des changements phonétiques, Berne.
- Passy, P. (1890): Etude sur les changements phonétiques, Paris: Librairie Firmin-Didot.
- Troubetzkoy, N.S. (1939): Grundzüge der Phonologie, Prague.

HIERARCHICAL RELATIONS AMONG PHONEMIC UNITS AS PHONOLOGICAL  
UNIVERSALS

Thomas V. Gamkrelidze, The Oriental Institute, Academy of Sciences,  
Georgian SSR, Tbilisi, USSR

After splitting the phoneme into its minimal components - distinctive features - and viewing it as a bundle of such features the question arises as to mutual compatibility of the features within the bundle and their relationship to one another on the axis of simultaneity.

It is the differing capacity of distinctive features for relating with one another into simultaneous combinations or "vertical sequences" that creates bundles of features differing in character and possessing a varying degree of "markedness", i.e., combinations of features characterized by commonness, naturalness, high degree of occurrence in the system ("unmarked") and less common, less natural combinations of features manifesting a lower degree of occurrence ("marked"), cf. Gamkrelidze (1975).

Depending on the varying capacity of distinctive features to combine with one another in a simultaneous bundle, it proves feasible to set up a gradation scale of "markedness" of simultaneous (vertical) combinations of features. Opposite extreme values on such a "scale of markedness" involve: (a) obligatory combination of the distinctive features on the axis of simultaneity, i.e., maximally "unmarked" combinations (as, e.g., combinations of features like [+syllabic, -nonsyllabic], [-syllabic, +nonsyllabic] or [discontinuous, dental], etc.) which are represented in any phonemic system being a constituent part of the phonemes entering the minimal phonemic inventory of language, and (b) noncombina- bility, mutual incompatibility of features potentially forming maximally "marked" combinations (e.g., the features of [glottalization] and [voice] or the features [nasal] and [fricative] that are incapable of combining into simultaneous bundles).

Between such extreme values of "markedness" are arranged all kinds of possible combinations of distinctive features with varying degrees of "markedness" - with a greater or lesser approximation to the extreme values reflecting the varying capacity of distinctive features to combine with one another in forming simultaneous bundles.

Such a "scale of markedness" of combinations of distinctive features must, in principle, be characterized by a fairly high degree of universality, for it reflects the capacity - common to human language - of definite phonetic and acoustic-articulatory properties to combine more or less freely and form simultaneous articulatory complexes. Definite phonetic features, owing to their acoustic-articulatory peculiarities, combine preferably with one another on the axis of simultaneity. "Marked" bundles of features reflect - in contrast to "unmarked" bundles - a limited capacity of definite phonetic features to join in simultaneous combinations, i.e., their lesser tendency to mutual combination. Hence such bundles represent less usual or less natural combinations of features, being placed on the "scale of markedness" closer to the maximal value of "markedness".

It is but natural to expect that such bundles (and correspondingly the phonemes represented by them) will be characterized by a lesser degree of actualization in language than will features which, in view of their acoustic and articulatory properties, combine easily with each other, representing natural or usual combinations of features. The former group of bundles of features (and correspondingly the phonemes represented by them) constitutes functionally weak units in the system, being characterized by a low degree of occurrence (frequency) and distributional limitations or being entirely absent in a number of languages, forming gaps in the paradigmatic system; the latter group of bundles is more common and natural and, in this sense, "unmarked", forming functionally strong units of the system and being characterized by a greater distributional freedom and a higher degree of occurrence (frequency) - some of them with a probability of occurrence equal to one (maximally "unmarked" combinations of features). Thus definite distinctive features combine with one another in simultaneous bundles in preference to other features, the combinations of which on the axis of simultaneity form more complex units in terms of articulation and perception. Being less optimal, such combinations are of a limited occurrence in the system, forming less natural phonemic units characterized by a lower frequency of occurrence and equalling zero in certain systems (yielding phonemic gaps in the paradigmatic pattern).

The phonemic units representing stable and "unmarked" bundles

of features in any linguistic system may be characterized as "dominant" as opposed to the less common and less natural (i.e., "marked") units of the system that may be styled "recessive". Thus, any two phonemic units opposed to each other in the paradigmatic system by the hierarchical relationship of "markedness" may be characterized as "dominant" vs. "recessive", while the relationship of "markedness" itself, implying a dependence between these units, may be restyled as the relation of "paradigmatic dominance". The terms are obviously borrowed from molecular biology, known for its ample use of linguistic vocabulary in application to the genetic code (cf. Jacob, 1977). Such a change of terms and the substitution of "dominant vs. recessive" for "unmarked vs. marked" seems to be expedient in view of the ambiguity of the traditional expression "markedness" and its still widespread use in the original sense of "merkmalhaltig/merkmallos" (as different from that of "common, natural" vs. "less common, less natural").

It is precisely the establishment of such universal patterns of compatibility of distinctive features into simultaneous bundles or into "vertical sequences", with determination of their opposite function of "dominance" in the paradigmatic system that appears to be one of the basic tasks of present-day typological phonology. This will help establish universally relevant hierarchical dependence between the correlative units of a phonological system and to identify the core of phonemic oppositions, a kind of deep phonological structure, that constitute the basis of the phonemic inventory of human language, invariant in relation to particular phonemic systems in synchrony and to possible phonemic transformations in diachrony.

In this respect correlations of stops and fricative phonemes in a phonemic system present a special interest. In particular, in the subsystem of stops the following hierarchical correlations of dominance may be established among the phonemic units of various series (cf. Melikishvili, 1970):

In systems with an opposition among stops differing on the feature "voice", the voiced labial /b/ is functionally stronger (dominant) as compared to the velar stop /g/. Stated otherwise, the feature "labiality" in a simultaneous combination with the feature "voice" yields a dominant bundle of features, making up the labial phoneme /b/, as different from the combination of the

features "voice" and "velarity" that yield a functionally weaker, less common and in this sense "recessive" voiced velar stop /g/. Inversely, in the class of voiceless stops it is precisely the velar /k/ that appears as a more natural, functionally stronger and dominant member of the paradigmatic opposition as compared to the labial /p/ serving as a functionally weaker, recessive unit. Thus "velarity" combined with "voicelessness" and "labiality" combined with "voice" form more natural and common bundles of features representing the dominant phonemes /k/ and /b/, whereas the combinations of "voicelessness" with "labiality" and of "voice" with "velarity" create functionally weak, recessive units /p/ and /g/, this being due to the acoustic-articulatory characteristics of the features involved.

Gaps in the paradigmatic system are distributed according to the established functional correlation of dominance of the phonemic units. Systems with gaps in the class of stops opposed according to "voice/voicelessness" assume in general the form as in (1-3):

(1) b -	(2) b p	(3) b -
d t	d t	d t
g k	- k	- k

The degree of recessiveness in the class of voiceless stops increases in accordance with the superposition on the bundle of the additional feature "aspiration" or "glottalization"; incidentally, "glottalization" appears as a feature of a higher degree of "recessiveness" than does "aspiration", so that the hierarchical sequence of increasing dependence in the class of unvoiced stops has the form: voiceless (plain) - aspirated - glottalized. Thus, the glottalized labial /p<sup>h</sup>/ appears in relation to the aspirated /p<sup>h</sup>/ as a recessive member of the opposition, whereas the aspirated /p<sup>h</sup>/ is recessive in relation to the voiceless plain phoneme /p/ (cf. Greenberg, 1970).

Gaps in the paradigmatic systems are represented in accordance with these correlations. The possible systems with gaps in the respective series of voiceless stops are given in (4) and (5):

(4) b p <sup>h</sup> -	(5) b - -
d t <sup>h</sup> t'	d t <sup>h</sup> t'
g k <sup>h</sup> k'	g k <sup>h</sup> k'

There appears to be a further dependence in the paradigmatic system between the subclass of stops and that of the corresponding fricative phonemes which manifest analogous correlations of dominance.

In the labial series the voiced fricative phoneme w/v/β emerges as the dominant member of the correlation, with the recessive voiceless unit /f/, whereas in the velar series the voiceless fricative /x/ functions as the dominant unit as opposed to the recessive voiced fricative /ɣ/, i.e., f → w/v/β, γ → x, and γ → w/v/β, f → x (where the arrow is directed from the recessive member of the opposition to the dominant one). Systems with gaps in the class of non-strident labial and velar fricatives with an opposition of "voice/voicelessness" assume in general the form as in (6-8):

(6) w/v f	(7) w/v -	(8) w/v -
- x	γ x	- x

The subsystem of fricatives appears in the paradigmatic system as a kind of substitute for the corresponding stops. In particular, the absence in the subsystem of stops of its functionally weak, recessive members (i.e., of the velar phoneme in the voiced series and/or the labial phoneme in the voiceless series) presupposes the presence in the paradigmatic system of the corresponding fricative phonemes (i.e., of the velar fricative in the voiced series, and/or the labial fricative in the voiceless series):  $\bar{g} \rightarrow \gamma$ ,  $\bar{p} \rightarrow f$ . Thus, the fricative phonemes /f/ and /ɣ/ and the dominant members implied by them, viz. w/v and /x/, respectively, emerge as substitutes for the corresponding stops /p/ and /g/, compensating, as it were, for their absence and thus establishing an equilibrium in the paradigmatic system. It may be asserted that the tendency to such an equilibrium in the system is due to the natural phonetic tendency to a symmetric filling of the three main articulatory zones - labial, dental, and velar - with sounds of consonantal articulation: stops or fricatives. If the system has the recessive stops /p/ and /g/, the presence of their substitutes in the form of the corresponding fricatives /f/ and /ɣ/ is optional. Such phonemes appear in the paradigmatic system as redundant consonantal elements, subject to diachronic changes.

Language systems evince a definite hierarchical order among diverse types of structural, in particular phonological, oppositions indicating the existence of a strict stratification of phonological values. It is in conformity with such universally valid correlations that diachronic phonemic transformations occur in a language. This gives a clue helping us to better understand language change in diachrony and to propose linguistically more realistic and plausible schemes of language reconstruction.

The classical Indo-European comparative grammar deals with a system of Proto-Indo-European stops that appears to be linguistically improbable and unrealistic since it runs counter to the typologically established phonological universals concerning the nature of the system of stops, with different phonemic series and a definite distribution of gaps. This necessitates a total revision of the traditionally postulated three-series-system of Proto-Indo-European stops - I: voiced II: voiced aspirates III: voiceless (with an absent or weakly represented voiced labial /b/) and its reinterpretation as I: glottalized II: voiced aspirates III: voiceless aspirates (with an absent, resp. weakly represented, glottalized labial /p'/), cf. Gamkrelidze-Ivanov (1973); Hopper (1973):

<u>Traditional System</u>				<u>Revised System</u>		
I	II	III		I	II	III
(b)	b <sup>h</sup>	p		(p')	b <sup>[h]</sup>	p <sup>[h]</sup>
d	d <sup>h</sup>	t		t'	d <sup>[h]</sup>	t <sup>[h]</sup>
g	g <sup>h</sup>	k	⇒	k'	g <sup>[h]</sup>	k <sup>[h]</sup>
.	.	.		.	.	.
.	.	.		.	.	.
.	.	.		.	.	.

Such a reinterpretation of the traditional system of Proto-Indo-European stops brings it in full conformity with typological evidence, both synchronic and diachronic, and allows to envisage a more realistic and linguistically plausible picture of Proto-Indo-European.

The evidence of the modern linguistic typology and the theory of language universals in effect necessitates a revision of the traditional schemes of the classical comparative linguistics by ad-

vancing new comparative-historical reconstruction.

This is one of the more practical aspects (finding its application in diachronic linguistics) of the modern linguistic typology and the theory of language universals.

#### References

- Gamkrelidze, Th. V. (1975): "On the correlation of stops and fricatives in a phonological system", *Lingua* 35.
- Gamkrelidze, Th. V. and V.V. Ivanov (1973): "Sprachtypologie und die Rekonstruktion der gemeinindogermanischen Verschlüsse", *Phonetica* 27.
- Greenberg, J. H. (1970): "Some generalizations concerning glottalic consonants, especially implosives", *IJAL* 36.
- Hopper, P. J. (1973): "Glottalized and murmured occlusives in Indo-European", *Glossa* 7.
- Jacob, F. (1977): "The linguistic model in biology", in *Roman Jakobson. Echoes of his Scholarship*, D. Armstrong and C.H. van Schooneveld (eds.), Lisse: The Peter de Ridder Press.
- Melikishvili, J. G. (1970): "Conditions of markedness for the features of voice, voicelessness, labiality, and velarity", *Matsne* 5.

## APPARITION ET DISPARITION DES OCCLUSIVES SONORES PRÉGLOTTALISÉES

André-Georges Haudricourt, Centre National de la Recherche Scientifique, 15, quai Anatole France, 75700 Paris, France.

Quelles sont les conditions linguistiques d'apparition des occlusives sonores préglottalisées? A la différence d'une occlusive sourde ou d'une spirante sonore, une occlusive sonore ne peut pas se prolonger indéfiniment. L'air qui passe à travers le larynx en produisant la sonorité est arrêté ensuite par l'occlusion buccale, de sorte qu'au bout d'un instant, la pression de l'air situé entre le larynx et l'occlusion buccale augmente et devient égale à celle de l'air des poumons et de la trachée-artère; de ce fait, l'air ne passe plus : la vibration laryngale s'arrête. Ainsi, la consonne sonore longue tend à s'assourdir. Pour maintenir la distinction pertinente, il faut diminuer la pression de l'air entre le larynx et l'occlusion buccale, c'est-à-dire fermer le larynx au début (préglottalisation), puis le faire descendre pendant la tenue de l'occlusion buccale. Le caractère injectif de la consonne n'en est que la conséquence, lorsque la désocclusion buccale se produit et que l'espace supraglottal a encore une pression inférieure à la pression atmosphérique.

Les langues indo-aryennes de la vallée de l'Indus, sous l'influence probable d'un substrat dravidien, ont transformé tous leurs groupes de consonnes en gémées; la pertinence phonologique d'une distinction entre simples et gémées, entre sourdes et sonores, avait un rendement important, et l'apparition des préglottalisées s'explique. Ces consonnes ne sont conservées actuellement qu'en sindhi, car c'est seulement dans cette langue que les sonores ordinaires sont réapparues assez tôt pour maintenir l'opposition<sup>1</sup>.

En Indonésie, on constate l'apparition de la préglottalisation comme réalisation d'occlusives sonores gémées dans certaines langues, tel le samal, parlé dans l'archipel Sulu des Philippines, ou en bougui, parlé à Sulawesi (Célèbes)<sup>2</sup>.

Les langues miao qui ont pénétré en Indochine depuis un siècle ont des occlusives latérales (mais pas de groupes de consonnes); or les langues indigènes d'Indochine n'ont pas ce type de consonnes. Dans le dialecte meo blanc, cette occlusive latérale a été considérée comme un groupe combinant occlusion et sonorité et est devenue une occlusive préglottalisée. Or, les préglottalisées

sont fréquentes dans les langues indigènes, ce qui a dû favoriser ce changement.

En vietnamien, les deux occlusives sourdes p et t sont devenues préglottalisées sonores au cours du Moyen Age, sans qu'aucune raison linguistique puisse être avancée. Il s'agit ici de causes ethnosociologiques; au cours du millénaire d'occupation chinoise, les anciennes préglottalisées austroasiatiques ont disparu (en devenant nasales) en vietnamien proprement dit, mais ont été conservées au voisinage (par exemple en müöng). Lorsque le Vietnam devint indépendant au X<sup>ème</sup> siècle, les préglottalisées réapparurent. Le même phénomène eut lieu en khmer : les p et t en contact avec la voyelle accentuée se sont préglottalisés, d'où la valeur donnée à ces lettres dans l'écriture thai dès le XII<sup>ème</sup> siècle.

Dans l'île de Hai-nan, deux langues introduites au Moyen Age – une langue thai, le bê, et un dialecte chinois min, le hainanais ou hoklo – ont subi cette même mutation des p-, t-<sup>3</sup>. Le même phénomène est signalé dans les dialectes yüe du sud-est du Guangxi<sup>4</sup>.

En résumé, en Indochine, les langues austroasiatiques et thai ont dû, au cours de leur évolution vers la monosyllabisation, engendrer des groupes initiaux de consonnes qui ont abouti linguistiquement à former des occlusives sonores préglottalisées (c'est ce qu'on constate dans une langue de Formose<sup>5</sup>), puis dans cette aire les langues qui venaient d'ailleurs en ignorant ces consonnes, ou celles qui historiquement les avaient perdues, les ont acquises par influence ethnosociologique. C'est le cas des langues karen : sgaw et pwo, langues tibéto-birmanes ayant pénétré dès le haut Moyen Age dans le domaine des langues austroasiatiques, et thai. Il y a passage de p-, t- aux sonores préglottalisées.

Actuellement, le thai de Bangkok a transformé ses préglottalisées en sonores ordinaires, peut-être au contact des langues européennes, mais les anciennes sonores historiques s'étant assourdies, la glottalisation n'avait plus de pertinence phonologique.

Par contre, lorsque les langues thai arrivent sur le domaine des langues tibéto-birmanes qui ignorent les préglottalisées, ces consonnes tendent à perdre leur occlusion en devenant des nasales préglottalisées (stade attesté par le ton, pour le khamti, le tay-nüa, le zhuang de Po-ai) qui sont maintenant des nasales ordinaires. Dans d'autres dialectes, elles deviennent des spirantes sonores (shan, tay-noir).

Le passage aux sonores ordinaires a dû se produire en khasi, langue austroasiatique isolée en domaine tibéto-birman.

Le penjabi, langue indo-aryenne voisine du sindhi, a dû passer par le même stade que celui-ci, car les occlusives sonores de cette langue sont liées au ton haut; elles ont été préglottalisées. Et lorsque les anciennes sonores se sont assourdiées (comme en khmer), la préglottalisation, n'étant plus pertinente, a pu disparaître.

#### Références bibliographiques

- (1) Haudricourt, A.-G. (1977): "La préglottalisation des occlusives sonores", Bulletin de la Société de Linguistique de Paris, 52, 1, pp. 313-317, Paris: Klincksieck.
- (2) Reid, L.A. (ed.) (1971): Philippine Minor Languages: Word Lists and Phonologies, p. 34, Oceanic Linguistics special publications n° 8 (256 p.), Honolulu: The University Press of Hawaii.  
 Reid, L.A. (1973): "Diachronic Typology of Philippine Vowel Systems" in Current Trends in Linguistics, 11, Diachronic, Areal and Typology, T.A. Sebeok (ed.), xii-604 p., The Hague: Mouton.
- Sirk, J.K. (1975): Bugijskij jazyk, p. 29, Moscou: édition "Nauka", 112 p.
- (3) Haudricourt, A.-G. (1959): "How History and Geography can explain certain phonetic developments", Yǔyán-yánjiu, 4, 81-86, Pékin.  
 Hagège C. et Haudricourt A.-G. (1978): La phonologie panchronique, Paris: Presses Universitaires de France.
- (4) Tsuji, N. (1977): "Murmured Initials in Yue Chinese and Proto-Yue Voiced Obstruents: the Case of Cenxi Dialect, Guangxi Province", Gengo Kenkyu, 72, 29-46, Tokyo.
- (5) Tsuchida, S. (1972): "The Origins of the Tsou Phonemes /b/ and /d/", Gengo Kenkyu, 62, 24-35, Tokyo.

TYPOLOGICAL UNIVERSALS, ASPIRATION, AND POST-NASAL STOPS

Robert K. Herbert, Department of Linguistics and Oriental and African Languages, Michigan State University, East Lansing, MI 48824, U.S.A.

Introduction

Probably the least marked type of consonant cluster found among the world's languages is Nasal + Oral Consonant (NC). The unmarked status of this sequence is demonstrated by a number of factors, including their occurrence in many languages otherwise characterized by CVCV structure. Perceptually, such a sequence is easily exploited since nasal consonants, although easily confused within the class, are quite distinct from oral consonants. The confusion within the class accounts, in part, for the fact that NC sequences are very frequently homorganic. Articulatorily, the sequence of gestures required to produce a NC cluster is relatively simple, involving only a raising of the velum for the sequence nasal plus voiced stop (ND).<sup>1</sup> For other types of NC clusters other gestures are necessary such as a cessation of vocal fold activity and a reduction in the degree of stricture. Further, the optimal opposition within NC sequences is demonstrated by its frequent exploitation in unit sound types, the so-called "half-nasal consonants", pre- and postnasalized consonants.<sup>2</sup>

The degree of articulatory and perceptual complexity is mirrored in the relative markedness of NC types. Thus, the least marked type of cluster is ND. Other types occur, even among the half-nasals, but these are less common<sup>3</sup> and many derivational processes, both synchronic and diachronic, conspire to produce NC inventories of the least marked type.

- 
- (1) The following symbol abbreviations will be used within the text: Nasal + Voiced Stop (ND), Nasal + Voiceless Stop (NT), Nasal + Voiced Fricative (NZ), etc. Other symbols employed have their standard phonetic values.
  - (2) The half-nasal consonants are distinguished from NC clusters by a number of factors, the most essential being that of duration. The two components of a half-nasal will exhibit the combined surface duration equivalent to a single consonant.
  - (3) This frequency is demonstrated in both cross-language frequency of occurrence and text frequency. In a 1000 phone count of a Rundi text, the following statistics were obtained: NC (30): ND (21), NT (4), NZ (4), NS (1).



A typological survey of the processes affecting either component of a NC sequence provides two inventories of process, one affecting the nasal and one the oral consonant. Among the former, only homorganicity assimilation is common whereas positional assimilation of the oral consonant is rare. Perhaps the most common process affecting the oral consonant is post-nasal voicing of voiceless consonants. In such a sequence, there are two primary motions which distinguish the two components: (1) raising of the velum, (2) cessation of vocal fold vibration. If the two are not coordinated the following sequences obtain: (a) NÇÇ, or (b) NŇÇ. In many languages the former tendency has been phonologized so that all post-nasal consonants are voiced.

Another common process is post-nasal hardening which, in conjunction with voicing, accounts for some of the many inventories containing only ND. Hardening actually involves two subtypes, but since many languages exhibit these in conjunction, it is perhaps best to view this situation as a continuum:

continuant → affricate → stop

In many cases, the hardening effect of nasals is evident even after the nasal is lost historically.

Other processes not of relevance to the present paper include post-nasal de-implosion (Shona /N+ð/→[mb]), ejectives (Zulu /N+ph/→[mp?]), etc. The situation with regard to aspiration of voiceless stops is problematic. On the one hand, some languages exhibit clear patterns demonstrating the loss of aspiration in this environment. However, other languages show aspiration developing in this context. Thus, there are conflicting tendencies which exist with regard to aspiration. This is not a felicitous situation since it is otherwise possible to determine a general direction of evolution. While changes of the sort NÇ → NŇ, NT → NTS occasionally occur, they are rare and other factors are found which explain these anomalous developments.

#### Loss of Aspiration

In Zulu, aspiration is lost in contact with nasal consonants. Doke (1926) reports the development of ejectives from aspirates in this context, but not all speakers exhibit this tendency. Aspirated clicks are replaced by simple nasal clicks when they are brought under nasal influence in Zulu whereas they merely lose

their aspiration in Xhosa. Tarascan (Foster 1969) has two series of underlying non-nasal obstruents /p t c č k/ and /p<sup>h</sup> t<sup>h</sup> c<sup>h</sup> k<sup>h</sup>/. In contact with nasal consonants, the plain consonants are voiced, and the aspirates become plain voiceless consonants, e.g. /N+p/ → [mb], /N+p<sup>h</sup>/ → [mp]. Devine (1974:19) notes that it may be best to regard this as a sliding scale of complexity and that the normal state for voiceless consonants in contact with preceding sonorants is unaspirated.

#### Development of Aspiration

In his useful survey of the noun class system of Bantu, Kadima (1969:63-5) notes that the most common developments of NT sequences are:

$$\begin{aligned} /N + p t k/ &\rightarrow [p t k] \\ &[p^h t^h k^h] \\ &[mp^h nt^h \eta k^h] \end{aligned}$$

Other developments also occur, e.g. [mb nd ŋg], [m̄ ŋ ŋ̄]. The present concern is with the development of aspiration. In Venda (Ziervogel and Dau 1961), Bantu nasal compounds develop as follow:<sup>4</sup>

*mb > mb	*mp > p <sup>h</sup>
*nd > nd	*nt > t <sup>h</sup>
*ŋg > ŋg	*ŋk > k <sup>h</sup>

When the simple stops are not under nasal influence, they undergo spirantization:

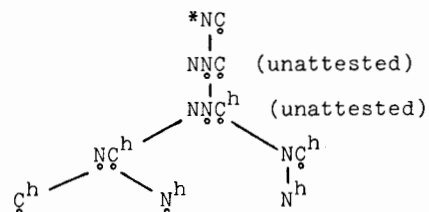
*p t k > φ r h
*b d g > β l Ø (j)

However, not all languages which develop aspiration exhibit a weakening of stops otherwise; it is therefore not possible to attribute aspiration to any general weakening process.

Hinnebusch (1975) attempted to reconstruct the phonetic processes in Swahili by which \*mp nt ŋk became p<sup>h</sup> t<sup>h</sup> k<sup>h</sup>. He proposed a two-stage process, the first of which is nasal devoicing, followed by deletion. It is proposed that native speakers reinterpreted the period of initial noisiness as post-aspiration rather

(4) However, the nasal is retained in both series with monosyllabic stems although it comprises a separate syllable: ŋk<sup>h</sup> 'large pot', nt<sup>h</sup> 'louse'. Further, the nasal is retained if it represents the first person singular object marker.

than preaspiration. The following is attributed to John Ohala:



Givon (1974) explains the development of aspiration by reference to three facts

- i) assimilatory devoicing of nasals before voiceless stops
- ii) voiceless nasals tend "to be perceived as breath"
- iii) voiceless stops tend to be universally aspirated

A perceptual confusion arises and there is a metathesis in which nasal breath is interpreted as post-aspiration. Ignoring for the moment the assertion that aspiration is the natural state for voiceless consonants universally, a universal of doubtful validity, the metathesis analysis seems plausible.

The two unattested stages above represent periods of variation before any phonetic tendency is phonologized. Wide variation in the realizations of NT sequences are found in many languages, e.g. in Malagasy /mp/ may be [mp, mp̚, ĥp, p, p̚, ph].

#### Further Development of the Aspirates

Aspirated stops derived in this manner are liable to other developments after the nasal has been lost. Frequently, they develop into fricatives or affricates. There is much comparative evidence to support this, e.g. Tswana *m̥haxo*, Pedi *mp̥<sup>h</sup>aYɔ*, Sutho *mofaɔ* 'provisions'. (Cf. also the development of postnasalized stops into aspirates and fricatives in many New Caledonian languages (Haudricourt 1964, 1971).) Languages frequently pass through an affricate stage before the fricative inventory is established. Hyman (1974) argues that even when there is no evidence for such a stage we may assume a "telescoping" of process. The important point here is to note that these developments occur only after the nasal has been lost; this explains why correspondences such as \*mp nt ŋk → f θ x do not violate the universal of hardening discussed above. Similarly, Sango \*ŋk > ŋx must have passed through an intermediate stage \*ŋkx (<\*ŋkh) which derives from aspiration being interpreted as a velar fricative due to the

acoustic similarities between the two. This seems plausible when we view the rest of the NT series: \*mp̚ > mh, \*nt > nh. In fact, the aspiration of velars in many languages is often phonetically [x], e.g. Scots Gaelic (Ternes 1973).

Another seemingly anomalous situation is presented by languages in which stops are voiced except after a nasal. This is certainly a preferred environment for voicing, yet there are correspondences such as Bulu:

*p > v	*mp > f
*t > l	*nt > t
*k > ø	*ŋk > k

It is necessary to explain the non-voicing of /f t k/ as resulting from previous aspiration, which prevented voicing in this position.

A complete inventory of processes affecting the derivation of NC sequences is beyond the scope of this short paper. Although there are relatively few processes which operate on ND sequences, really only simplification in favor of the oral or nasal consonant, a number of processes conspire to produce NC inventories which include only ND sequences. These include both direct and indirect processes, i.e. those which change feature specifications and those which eliminate one component of the sequence. Apart from the universal primacy of ND sequences, there may be language-specific variation in terms of the relative weightings of other types, e.g. NT, NZ.

#### Typology and Reconstruction

Part of the value of surveys of evolutionary processes is that they serve as useful tools in diachronic linguistics. This idea is far from novel. Jakobson (1958) noted that such studies form the touchstone of validity for all reconstructed systems. The interaction of processes of change as well as the directionality of change itself often provide insight into problems of reconstruction. Studies of this sort point not only backwards to possible sources of origin, but also forwards to future directions of possible change.

Bennett (1967), in discussing the voicing of post-nasal stops in several Eastern Bantu languages, reconstructs the phonetics of change as:

\*mp > \*mɸ > \*mɸ > mb

\*nt > \*nθ > \*nð > nd

Nasality is lost in certain cases and \*mp > ɸ or f. However, there are several serious problems with the proposed reconstruction. Specifically, the change \*mp > mɸ is unlikely to the extent that consonants tend to harden in this environment. In fact, the sequences [mɸ mɸ nθ nð] are all uncommon. Although [mɸ] occurs, it always represents /mɸ/, never /mb/, and the more common realizations of such a sequence are [mb, ɸ, b]. Ladefoged (1968:47) reports the existence of [nθ] in Sherbro, a surprising fact since Sherbro also exhibits /f v s/, none of which appear after a nasal. Kamba exhibits [nð]. On the whole, however, this is a restricted class of sounds.

Further, the fact that intervocalic voiceless stops lenite cannot be cited as evidence that post-nasal stops behave similarly. There are numerous examples where the two develop differently. For example, Londo \*mp nt ŋk > p t k whereas p t k > ɸ t x. In Mbɔle, \*mp nt ŋk > f t k and \*p t k > ɸ t ø. A crucial fact in cases exhibiting the development of a fricative from a voiceless stop is that nasality is lost. In such a case, intermediate stages are attested elsewhere, e.g. Lwena \*mp nt ŋk > p<sup>h</sup> t<sup>h</sup> k<sup>h</sup> and p t k > h t k. Also, the existence of nasal and fricative series generally implies the existence of nasal and stop series, which condition is not met by Bennett's system. Thus, the proposed reconstructed chronology cannot be accepted, especially in view of the frequency and naturalness of the process whereby consonants are voiced after a nasal consonant.<sup>5</sup> The point here is that although it is necessary to make inferences about the phonetics of prehistory, these inferences must be solidly grounded in a theory of universal processes and phonetics. There are definite limitations to be placed upon the importance attached to such studies for other purposes,

(5) Cases such as Makua \*mb nd ŋg > p t k must involve two distinct stages: (1) nasal loss, (2) later devoicing. There is no neutralization of NC series since \*mp nt ŋk > p<sup>h</sup> t<sup>h</sup> k<sup>h</sup>. One step neutralizations of NC series always favor the voiced series, e.g. Yao \*mp, mb > mb; \*nt, nd > nd; \*ŋk, ŋg > ŋg.

e.g., genetic classification, linguistic subgroupings, etc.

#### Conclusion

This brief paper has attempted to demonstrate how various claims made by Jakobson, Greenberg, and others may be applied to the study of NC sequences. This included an examination of the relationship between synchronic universals and diachronic processes and between typology and universals. Greenberg (1970a:61) points out that the former follows logically from the fact that no change can produce a synchronically unlawful state and that all states are the outcome of diachronic processes. The distinction between state and process is an important one. The general direction of NC evolution toward the least marked ND sequence again supports the generalization that diachronic process explains frequency in phonology. The predictive power of typological studies demonstrates this complex interaction between the shape and patterning of phonological systems.

#### References

- Bennett, P. (1967): "Dahl's Law and Thagicũ" African Language Studies 8, 127-59.
- Devine, A. (1974): "Aspiration, universals, and the study of dead languages", Working Papers in Language Universals 15, 1-24.
- Doke, C. (1926): The Phonetics of the Zulu Language, Bantu Studies, Special Number.
- Foster, M. (1969): The Tarascan Language, Berkeley: UCPL 56.
- Givon, T. (1974): "Rule un-ordering: generalization and degeneralization in phonology", Papers from the Parasession on Natural Phonology, 103-15, Chicago: Chicago Linguistic Society.
- Greenberg, J. (1969): "Some methods of dynamic comparison", in Substance and Structure of Language, J. Puhvel (ed.), 147-203, Berkeley: University of California Press.
- Greenberg, J. (1970a): "Language Universals", in Current Trends in Linguistics, T. Sebeok (ed.), 3, 61-112. The Hague: Mouton.
- Greenberg, J. (1970b): "The role of typology in the development of a scientific linguistics", in Theoretical Problems of Typology and the Northern Eurasian Languages, L. Dezső and P. Hajdú (eds.), 11-24, Bucharest: Akadémiai Kiado.
- Guthrie, M. (1967-70): Comparative Bantu, 4 vols., Farnborough: Gregg International Publishers.

26 SYMPOSIUM No. 1

- Haudricourt, A. (1964): "Les consonnes postnasalisées en Nouvelle Calédonie", Proc. Ling. 9, 460-61, The Hague: Mouton.
- Haudricourt, A. (1971): "Consonnes nasales et demi-nasales dans l'évolution des systèmes phonologiques", Proc. Ling. 10, 4, 105-8, Bucharest: l'Académie de la République.
- Herbert, R. (1977): Language Universals, Markedness Theory, and Natural Phonetic Processes: The Interactions of Nasal and Oral Consonants, Unpublished Ph.D. dissertation, Ohio State University.
- Hinnebusch, T. (1975): "A reconstructed chronology of loss: Swahili class 9/10", in Proc. of the Sixth Conference on African Linguistics, R. Herbert (ed.), 32-41, Columbus: OSUWPL 20.
- Hyman, L. (1974): "Contributions of African linguistics to phonological theory", Fifth Conference on African Linguistics, Stanford University. Ditto.
- Jakobson, R. (1958): "What can typological studies contribute to historical comparative linguistics?", Proc. Ling. 8, 17-35, Oslo: Oslo University Press.
- Kadima, M. (1969): Le système des classes en bantou, Leuven: Vander.
- Ternes, E. (1973): The Phonemic Analysis of Scottish Gaelic. Hamburg: Helmut Buske Verlag.
- Ziervogel, D. and R. Dau (1961): A Handbook of the Venda Language, Pretoria: University of South Africa.

## UNIVERSALS OF VOWEL SYSTEMS: THE CASE OF CENTRALIZED VOWELS

Jean-Marie Hombert, Linguistics,  
University of California, Santa Barbara, USA 93106

This paper attempts to explain why centralized vowels (i.e. vowels which are not located on the periphery of the vowel space) are relatively less common than peripheral vowels.

1. Surveys of phonemic systems, phonetic universals and "exotic" languages.

If one is interested in discovering phonetic universals some of the most fruitful places to search for potential universals are large scale surveys of phonetic and phonemic inventories. Despite the criticism leveled against these surveys it is our belief that such surveys are useful in that asymmetries or systematic gaps in these inventories may reveal in their explanation universal phonetic processes. Once such a potential universal or universal tendency has been uncovered each language exhibiting this process should be reexamined through careful study of available sources, consideration of possible reinterpretations of the data, and when possible, accurate phonetic data should be obtained.

Until very recently the bulk of available phonetic data, especially perceptual data, has come from a handful of languages. Due to the availability of phonetic equipment and presence of research groups located in the countries where these languages are spoken available phonetic data has been largely limited to Danish, Dutch, English, French, German, Japanese and Swedish. It is clear that if we are to understand universal phonetic processes, our data base must be extended to include more "exotic" languages.

Most perceptual data has been gathered from experiments conducted under laboratory conditions using linguistically sophisticated subjects. Obviously if we are to gather similar data from languages spoken in areas remote from laboratory facilities, it is necessary to design techniques of data gathering suitable for use in the field with linguistically naive subjects. In Section 3 one such design will be discussed.

2. The case of centralized vowels.

It is clear from surveys of vowel systems that centralized vowels are less commonly found than peripheral ones. In the case of languages which do have centralized vowels it is not rare that different sources will vary in the treatment of such vowels by

attributing to a given vowel different phonetic qualities. These variations suggest that either these vowels are more prone to historical change or are more difficult to identify correctly by the investigator. It appears, then, from these surveys that non-peripheral vowels, that is, vowels which in acoustic terms have a second formant of approximately 1200-1700 Hz, are rare and that they are more subject to change than peripheral vowels.

In Section 3 we will use data from a perceptual experiment carried out on the Grassfield Bantu languages of Cameroon. Because of space constraints in this paper, we will use only data from one speaker of the Fe?fe? language<sup>1</sup> to suggest possible explanations for the rarity as well as instability of non-peripheral vowels.

### 3. Experimental paradigm

Fe?fe? contains eight long vowels in open syllables. These vowels are [i, e, a, v, o, u, ɛ, ə]. A word list consisting of eight meaningful Fe?fe? words contrasting these eight vowels was elicited from native Fe?fe? speakers. The Fe?fe? speakers were asked to read these eight words which were listed five times each, in random order. After the repetition of each word, the final sound of the word, that is the vowel, was repeated once. Both the vowels of the meaningful words and the vowels in isolation were subsequently analyzed.

Subjects were then asked to listen to 53 synthetic vowel stimuli, each presented five times in random order. After the presentation of each stimulus the subjects were instructed to point out which Fe?fe? word in the eight-word list that they had previously read, contained the same "final sound", i.e. vowel, as the stimulus. Subjects had the option to claim that some of the stimuli did not sound like any of the eight Fe?fe? words. The 53 synthetic stimuli were selected to maximally cover the vowel space; F1 was varied between 250 Hz-750 Hz, F2 between 650 Hz-2350 Hz and F3 between 2300 Hz-3100 Hz. This task was designed so that native speakers would divide the vowel space according to their own vowel systems.<sup>2</sup>

### 4. Results

The results of the acoustic analysis and of the perceptual

- 
- (1) For more data and a more complete description of the experimental paradigm, see Hombert (in preparation).
  - (2) It should be noticed that this method does not allow study of diphthongs since all stimuli used have steady state formant frequencies.

experiment for one Fe?fe? speaker are presented in Figure 1 and Figure 2 respectively. Since F3 values are not relevant for the point that we want to make here the data are presented in an F1 x F2 space. Each vowel indicated in Figure 1 is the average of five measurements. The spectra were computed 100 msec. after vowel onset using LPC analysis. The phonetic symbols appearing in Figure 2 indicate that at least four times out of five this stimulus was identified by the Fe?fe? speaker as the same vowel.

We will consider the two vowels [a] and [ə]. Two unexpected results emerge from the data:

1. When comparing acoustic and perceptual data it is not surprising to find that the stimulus with F1 at 750 Hz and F2 at 1250 Hz is identified as the vowel [a] since a vowel with such a formant structure could have been produced by a Fe?fe? speaker with a larger vocal tract size than the speaker considered here. What is surprising, though, is that the stimulus with the formant structure F1 at 750 Hz and F2 at 850 Hz was also identified as [a]. These results are even more surprising when one considers that the intermediate stimulus (750 Hz - 1050 Hz) was identified as [v]. It is likely that in the case of the stimulus with F1 at 750 Hz and F2 at 850 Hz the two formant peaks were perceived as one formant peak, that is as F1. One thing remains to be explained: in the acoustic data, the Fe?fe? vowel [a] has a peak around 1600 Hz but the stimuli with F1 at 750 Hz and F2 at 850 Hz does not have a peak in this frequency region. Let us just say for the moment that the saliency of the peak at 1600 Hz seems to be perceptually secondary.
2. Two stimuli (F1 at 350 Hz, F2 at 1500 Hz and F1 at 450 Hz, F2 at 1500 Hz) are identified as [ə], which is what we would expect considering the location of [ə] in Figure 1. However the identification of the stimulus with F1 at 450 Hz and F2 at 650 Hz with [ə] comes as a surprise. Notice that F1 and F2 are also close to each other for this last stimulus, which could have lead to the perception of them as one peak corresponding to the first formant. But notice also that this stimulus does not have a peak around 1500 Hz. As in the case of the vowel [a] it appears that the perceptual saliency of the peak around 1500 Hz did not play a major role in the identification of the [ə].

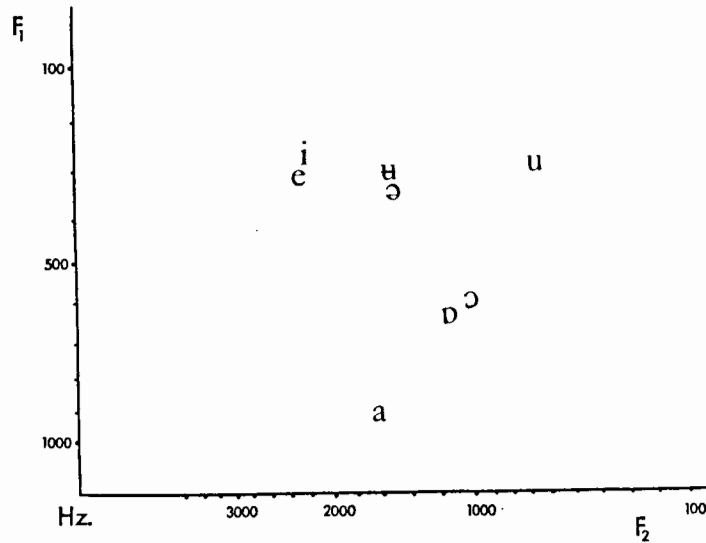


Figure 1. Acoustic data: the Fe?fe? vowel system, (one speaker, average of five measurements).

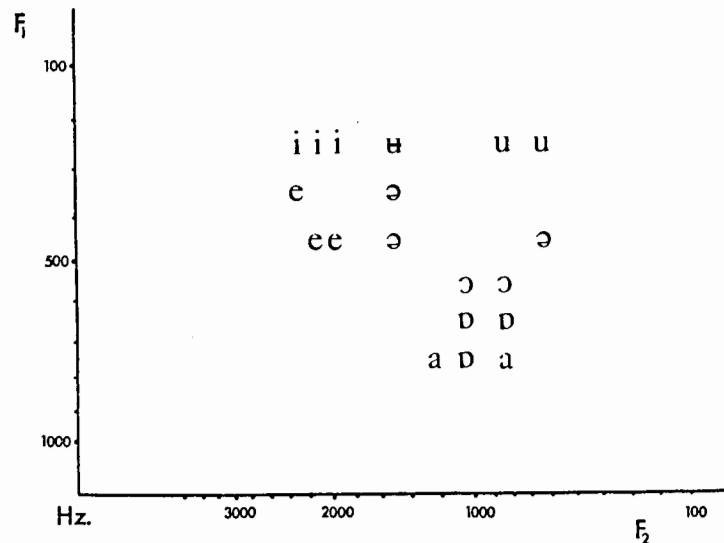


Figure 2. Perceptual data: only stimuli for which the Fe?fe? subject gave at least four out of five identical responses are presented on this graph.

5. Discussion<sup>3</sup>

Two possible explanations to account for the lack of saliency of formant peaks around 1500 Hz are being explored now.

1. Spectrum-based representation of vowels.

Our results would be compatible with a mechanism of vowel perception which looks for certain amounts of energy within frequency regions rather than formant peaks. In the cases which we discussed in the previous section, the unexpected vowel identification happened with stimuli which had their first and second formants very close to each other. In such cases the closeness of the first two peaks leads to an increase in amplitude of the spectrum. This increased amplitude may have created sufficient energy in the 1500 Hz region to lead to these "perceptual mistakes".

2. Place vs. periodicity mechanisms.

Pitch is processed by different mechanisms depending upon its frequency region. The boundary between these two mechanisms (place vs. periodicity) is not well defined. It is possible that for some subjects a defective overlap between these two mechanisms in the 1500 Hz region could create the perceptual mistakes presented in Section 4.

6. Implications

The explanation generally provided to account for the relative scarcity of non-peripheral vowels is based on the principle of maximum perceptual distance presented by Liljencrants and Lindblom (1972). Our results suggest a different explanation - non-peripheral vowels are avoided because one of their components (F2) is located in a relatively less salient perceptual zone. If this is the case we can now understand why processes leading to vowel centralization (vowel nasalization, rounding of front vowels, unrounding of back vowels) are relatively uncommon.

Finally we should point out that "perceptual mistakes" such as the ones reported in Section 4 were found in approximately one out of five subjects, with the "mistake" being consistently made by the one subject. These results would be consistent with a theory of sound change which claims that sound changes are initiated by a minority of speakers.

(3) The reason why previous experiments on vowel perception did not uncover this problem may be due to the nature of the experimental paradigm as well as the range of stimuli used in this experiment.

- (vii)  $\left\{ \begin{array}{l} a \rightarrow \text{ɔ/w} \text{ \_\_} \\ \emptyset \rightarrow \text{æ/} \text{ \_\_} r \end{array} \right.$  (was, swan, quarrel; Middle English).  
 ([ $\emptyset$ :va] vs [ $\text{æ}$ :ra]; Swedish).
- (viii)  $\left\{ \begin{array}{l} x \rightarrow \left\{ \begin{array}{l} \text{ç / +front V \_\_} \\ x / \text{elsewhere} \end{array} \right. \\ /h/ \text{ realizations of Japanese cf. (v) above.} \end{array} \right.$  ([l<sub>1</sub>çt] vs [axt]; German).
- (ix)  $\left\{ \begin{array}{l} n \rightarrow m / b \text{ \_\_} \\ /n/ \text{ realizations of Swedish cf. (iii) above.} \end{array} \right.$  ([ha:bm] (haben); German).
- (x)  $\left\{ \begin{array}{l} r \rightarrow \text{ʀ} / \left[ \begin{array}{l} \text{-voic} \\ \text{-son} \end{array} \right] \text{ \_\_} \\ \text{ʒ} \rightarrow \text{ʒ} / \text{ \_\_} \text{[-voic]} \\ k \rightarrow \text{k} / \text{ \_\_} \text{[+voic]} \end{array} \right.$  ((try, cry, pry; English).  
 ([neʒʒfɔ̃dy] French  
 [saʔdɔ̃ʀ] French)

The above examples of pro- and regressive assimilations suggest that assimilation be hypothetically described as a reduction of articulatory distance in articulatory space. Do they imply a syntagmatic pronounceability condition, favoring a reduction of the physiological equivalent of a power constraint, mechanical work (force x distance)/time (a LESS EFFORT principle)? Can at least some phonological facts be interpreted as cases of contrast-preserving articulatory simplifications? What is their behavioral origin?

### 3. Speech - a Physiological Pianissimo.

3.1 The question also arises whether spoken language underexploits the degrees of freedom that in principle the anatomy and physiology of speech production make available. Seen against the full range of capabilities, speech gestures, like many other skilled movements, appear to be physiologically "streamlined" both as regards muscle recruitment and force levels (cf. jaw closure as a speech gesture and in mastication, speech breathing vs respiration in general, articulatory gestures vs swallowing etc.). Extreme displacements of articulatory organs do not occur (PIKE 1943, 150) although such configurations are available and yield acoustically equivalent results (evidence from non-speech: body-arm, eye-head coordination; and from speech: lip/tongue-mandible and tongue blade-tongue body coordination (LINDBLOM et al 1974)). Do we in these circumstances see the operation of an economy of effort principle? A principle that we should invoke to explain how and why speech and non-speech sounds differ

and to account for certain phonological regularities as well as the instances of hypo-articulation (reductions, ellipses, co-articulations etc.) in spontaneous speech. "Today's allophonic variation leads to tomorrow's sound change..." OHALA (1979).

### 3.2 Pronounceability and Syllable Structure.

FIG. 1 shows average measurements of jaw positions for Swedish apical consonants in the environment [a'Ca:]. The production of these consonants permits a variable influence of the open jaw positions of the vowels. Thus the dimension of jaw opening reveals one aspect of their "willingness" to coarticulate. It is of considerable interest to see that this measure correlates well with their universally favored position in initial and final phonotactic structures (ELERT 1970). If the present observations are generalized, they imply that the phonetic structure of clusters can be explained at least in part with reference to ease of co-articulation (ELERT 1970, BRODDA 1972).

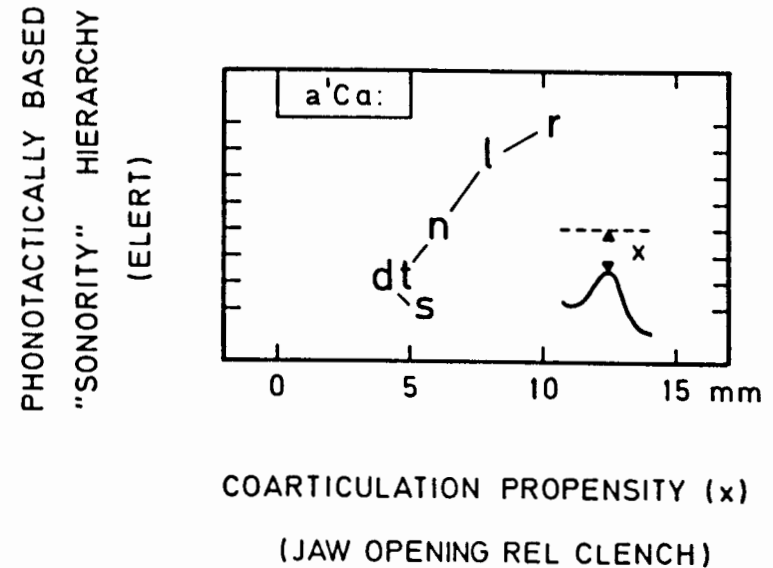


FIG. 1



## 4. The Distinctiveness "Conspiracy".

4.1 Language structure exhibits redundancy at all levels.4.2 Speech generation is an output-oriented process: The reference input to the speech control system is specified in terms of a desired output. The dimensions of the target specifications are sensory, primarily auditory. Evidence supporting the primacy of auditory targeting comes from work on compensatory articulation, speech development and the psychological reality of phonological structure (LINDBLOM et al to appear, LINELL 1974).4.3 Speech understanding is an active (top-down or conceptually driven) process. (Cf. the demonstrations of context-sensitive processing, resistance to signal degradation, phonemic restoration, verbal transformation etc.)4.4 The speech system may possess specialized mechanisms that contribute towards enhancing the distinctiveness of stimulus cues. Examples of such hypothetical mechanisms are "feature detectors" in speech perception. Specialization of speech production has been suggested in the case of the phylogenetic development of the human supralaryngeal vocal tract whose shape LIEBERMAN (1973) interprets as a primarily speech-related adaptation increasing the acoustic space available for speech sounds.4.5 Phonetic targets are selected so as to retain acoustic stability in the face of articulatory imprecision (STEVENS 1972).

The properties listed in 4.1 through 4.3, do they have a common origin in a basic principle of language design viz., the DISTINCTIVENESS CONDITION: different meanings sound different? The preservation of meaning across encoding and decoding seems to be favored by redundancy, output-oriented and active processing (rather than by lack of redundancy, exclusively input-oriented encoding and purely passive decoding strategies). Thus the question arises whether these at first seemingly unrelated attributes form an evolutionary "conspiracy". Do they constitute three different ways of coping with a physical signal which is inevitably going to be noisy, variable and ambiguous? 4.4 and 4.5 could offer related advantages. What is the behavioral origin of the distinctiveness condition?

## 5. Speech Development.

5.1 Imperfect learning: Can perceptual similarity and articulatory reinterpretation serve as a source of phonological innova-

tion (cf. JONASSON (1971))? Many sound substitutions in children's speech appear compatible with this interpretation:  $\theta \rightarrow f$ ,  $\lambda \rightarrow w$  cf. 2.1. The child is a cognitive and phonetic bottle-neck through which language must pass. Does the process of acquisition leave its imprints on language structure?

5.2 Selection of the fittest: A speech community may use in free variation several realizations of a given form. The set of fricatives may contain /f, s, ʃ, ç/ and /h/ with the /ʃ/ produced as [ʃ] and [ʃ̥] (cf. Swedish). The distinctiveness principle favors [ʃ] which contrasts better with [ç] than [ʃ̥]. The lower confusion risk of the pair [ʃ] / [ç] promotes its reception and learning by the child. There is in this case thus a behavioral rather than teleological motivation for the distinctiveness condition. If sound patterns show evidence of perceptual differentiation, is communicative "selection of the fittest" among several competing forms one of the evolutionary mechanisms? Selection occasionally occurs from a rich variety of hypo- as well as hyper-articulated forms (STAMPE 1972). Is hyperarticulation another behavioral source of distinctiveness?

## 6. Non-Phonetic Origins of Sound Patterns: Social Biasing.

Selection of speech forms is influenced not only by production and perception factors. Phonological contrasts vary as a function of social variables (prestige, age, class, sex, style etc.). Does the interaction of the sometimes conflicting requirements of social and phonetic factors account for the fact that there is no evidence (GREENBERG 1959) that language change leads to more efficient linguistic systems? Is local rather than global phonetic evaluation of systems (KIPARSKY 1975) another reason why languages do not seem to be converging toward a single optimum equilibrium?

The emergence of a phonological system can be simulated on the basis of current models of production and perception. FIG. 2 shows some computational results obtained by an application of

$$\sum_{i=2}^n \sum_{j=1}^{i-1} T_{ij}(t) \cdot L_{ij}(t) \cdot S_{ij}(t) < \text{CONSTANT} \quad (1)$$

where  $n$  is the size of a universal inventory of segments,  $T_{ij}$  represents a (time-varying) talker-dependent measure of evaluation for a given contrast (pronounceability condition),  $L_{ij}$

refers to a listener-dependent evaluation (distinctiveness condition), and  $S_{ij}$  reflects the balance between social and phonetic factors. FIG. 2 illustrates the

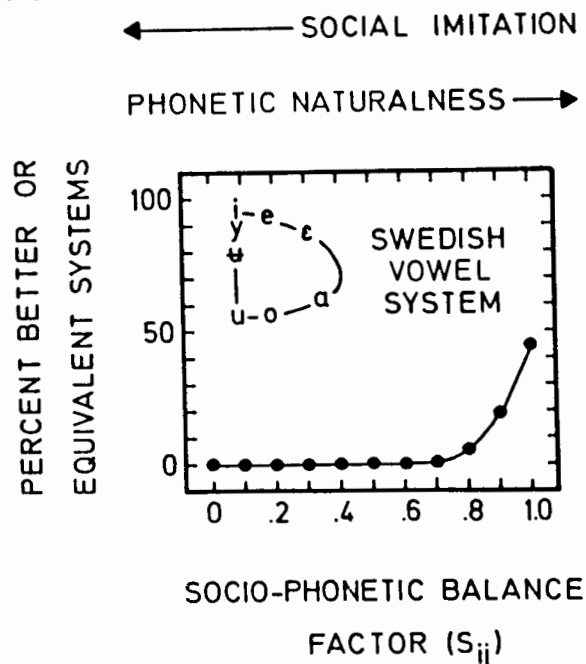


FIG. 2

interaction between the criteria of distinctiveness and social imitation in deriving the Swedish vowel system from a larger set of universal vowel types (represented in terms of canonical auditory patterns). The socio-phonetic balance varies from zero ("social imitation" dominates) to unity (natural phonetic factors, T and L, dominate). It is applied to the contrasts of Swedish with the values shown. For non-Swedish contrasts  $S=1$ . Apparently there are many systems (out of a total of 92378) that meet our present criterion of distinctiveness equally well or better. If we had reason to believe that the role of natural phonetic factors in the genesis of the Swedish vowels was correctly and exhaustively reflected in our calculations we would conclude that social factors are quite important in their development. We don't. A great deal of work on phonetic naturalness remains to be done before any safe conclusions can be drawn.

However, we believe that the approach will be useful in studying phonological contrasts particularly in child language and cross-linguistically.

#### 7. A "Darwinian" Theory of Phonological Universals.

Suppose that we answer all the questions of the preceding discussion in the affirmative. We accept as our null hypotheses the assumptions that learnability, pronounceability and perceptibility conditions can account for differences between speech and non-speech sounds, that discreteness reflects the operation of memory, learning and decoding mechanisms, that sound changes are influenced by social variables and shaped by demands for perceptual efficiency and convenience of production, and that the origin of such demands is prosaically behavioral rather than mysteriously teleological. Such acceptance boils down to the idea that phonological structure arises both phylogenetically and ontogenetically by "natural selection" of sound patterns from sources of phonetic variation. Language structure emerges in response to the biological and social conditions of language use. Natural selection is based on the communicative (perceptual as well as social) value of contrasts and "phonetic variation" is defined with respect to possible segment, possible sequence and their possible variation. According to this "Darwinian" theory, phonological universals will be explained with reference to how speech is acquired, produced and understood, or rather in terms of our models of these processes.

This conclusion may seem uncontroversial. However, a truly quantitative and explanatory theory along these lines is not likely to appear until we learn to recognize its full intellectual, educational and administrative implications for how linguistics should be done. Language is the way it is partly because of our brains, ears, mouths as well as our minds. In this sense linguistics is a natural science. Phonetics can contribute by formulating its behavioral explanans principles.

#### 8. References.

- BRODDA, B. (1973): "Naturlig Fonotax", unpubl. manuscript, Stockholm University.
- CHAFE, W.L. (1970): Meaning and Structure of Language, Chicago and London: The University of Chicago Press.
- ELERT, C.C. (1970): Ljud och Ord i Svenskan, Stockholm: Almqvist & Wiksell.

- GREENBERG, J.H. (1959): "Language and Evolution", in MEGGERS, B.J. (ed.): Evolution and Anthropology: A Centennial Appraisal, pp. 61-75.
- JONASSON, J. (1972): "Perceptual Factors in Phonology", in RIGAULT, A. & CHARBONNEAU, R. (eds.): Proceedings in the Seventh International Congress of Phonetic Sciences, pp. 1127-1131, The Hague: Mouton.
- KIPARSKY, P. (1972): "Explanation in Phonology", in PETERS, S. (ed.): Goals of Linguistic Theory, pp. 189-227.
- KIPARSKY, P. (1975): "Comments on the Role of Phonology in Language", in KAVANAGH, J.F. and CUTTING, J.E. (eds.): The Role of Speech in Language, pp. 271-280.
- LIEBERMAN, P. (1973): "On the Evolution of Language: A Unified View", Cognition 2 (1), pp. 59-94.
- LINDBLOM, B., PAULI, S. and SUNDBERG, J. (1974): "Modeling Coarticulation in Apical Stops", in FANT, G.: Speech Communication, vol. 1, pp. 87-94, Almqvist & Wiksell Int.
- LINDBLOM, B., LUBKER, J. and GAY, T. (in press): "Formant Frequencies of Some Fixed-Mandible Vowels and a Model of Speech Motor Programming by Predictive Simulation", J. Phonetics.
- LINELL, P. (1974): "Problems of Psychological Reality in Generative Phonology: A Critical Assessment", Reports from Uppsala University Department of Linguistics nr 4.
- MANDELBROT, B. (1954): "Structure Formelle des Langues et Communication", Word 10, pp. 1-27.
- MILLER, G.A. (1956): "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information", Psychological Review 63, pp. 81-97.
- OHALA, J.J. (1979): "The Contribution of Acoustic Phonetics to Phonology", to be published in LINDBLOM, B. and ÖHMAN, S. (eds.): Frontiers of Speech Communication Research, London: Academic Press.
- PIKE, K.L. (1943): Phonetics, Ann Arbor: The University of Michigan Press.
- STAMPE, D. (1972): "On the Natural History of Diphthongs", Papers from the 8th Regional Meeting, Chicago Linguistic Society, pp. 578-590.
- STEVENS, K.N. (1972): "The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data", in DAVID, E.E. and DENES, P.B. (eds.): Human Communication: A Unified View, New York: McGraw Hill.

## UNIVERSALS OF LABIAL VELARS AND DE SAUSSURE'S CHESS ANALOGY

John J. Ohala, Phonology Laboratory, Department of Linguistics,  
University of California, Berkeley, California, U.S.A.

In the Cours de linguistique generale de Saussure compares language to a chess game and the units of the linguistic code to the individual chess pieces. He remarks that

If I use ivory chessmen instead of wooden ones, the change has no effect on the system ... The respective value of the pieces depends on their position on the chessboard just as each linguistic term derives its value from its opposition to all the other terms ... [A single] move has a repercussion on the whole system [1916 (1966:22; 88-89)].

The choice of the chess analogy was a brilliant piece of exposition. Justifiably, it is frequently cited, especially by teachers in linguistic courses, and has become one of the favorite images of the structuralist basis of linguistic analysis.

The structuralist approach in phonology means analyzing a given problem by taking the whole phonological system into consideration, e.g. all the phonemic oppositions, especially those which are symmetrical or asymmetrical, the functional load of the sounds involved, etc. It focuses, therefore, on system-internal relations between the 'pieces', i.e., the speech sounds. For example, the structuralist account of the introduction of [ʒ] into English would point out that it filled what was up to that time a gap in the English fricative system:

f	θ	s	ʃ
v	ð	z	ʒ

Generative phonology, a recent offshoot of structural linguistics, also focuses on system-internal relations between the 'pieces' although in this case the pieces are the rules of grammar and the entities which make up the lexicon.

In fact, almost any post-Saussurean "school" of phonology one might cite, e.g., the Prague school, glossematics, functional phonology, natural generative phonology, etc. -- all have adopted the structuralist method of looking within the system for the solution to their problems. Occasional explorations outside the system -- into anatomy, physiology, physics, psychology, etc. have never been

pursued seriously or intently.<sup>1</sup>

I would maintain that this emphasis on system-internal relations in phonology is counter-productive. This point is especially evident when we examine and seek explanations for phonological universals. We frequently find speech sounds behaving in very similar ways across languages even though those languages exhibit remarkably varied structure. The phonological behavior of labial velars, i.e., [u, w, m,  $\widehat{kp}$ ,  $\widehat{gb}$ ] etc. illustrates this rather dramatically.<sup>2</sup>

It has been claimed by generative phonologists that labial velars, although possessing two more or less equal constrictions, labial and velar, nevertheless, must be represented at the underlying (lexical) level as having only one primary articulation -- the other constriction being relegated to a secondary articulation (Chomsky and Halle 1968, Anderson 1976). The phonological behavior of a segment is supposedly a function of its underlying representation, not its surface phonetic character. Thus Anderson, in reviewing a number of West African languages, argues that Temne which has a /k/ but no /g/, must classify its / $\widehat{gb}$ / as a velar, filling the gap in the voiced velar stop position. Similarly he argues that since Nkonya has both /k/ and /k<sup>w</sup>/, the second sound thus preempting the classification: 'primary articulation: velar; secondary articulation: labial', the sound / $\widehat{kp}$ / in that language must be primarily a labial with a secondary velar articulation. Efik, he notes, not only has a /k/ vs. /k<sup>w</sup>/ contrast but also lacks a /p/, so it has two reasons for classifying its / $\widehat{kp}$ / as a labial.

One of the problems with such structural or functional accounts of phonological facts is that they attach undue significance to sound patterns which may commonly arise due to chance or at least due to factors unrelated to the particular phenomena under investigation. Attention to phonological universals would be some insurance against this problem. As it happens, /p/ and /g/ are often missing from languages' stop inventories (Gamkrelidze 1975, Sherman

(1) Notable exceptions, however, are the fields of sociology, cultural history, and anthropology, which have been pursued seriously by many phonologists with structuralist orientation.

(2) The research on labial velars was done in collaboration with James Lorentz and published in Ohala and Lorentz (1977). Limitations of space prevent extensive documentation of the sound patterns discussed; however, the article cited may be consulted for numerous cross-linguistic examples.

1975). Moreover, there are many languages in West Africa that have / $\widehat{kp}$ / and/or / $\widehat{gb}$ / (Ladefoged 1964). Why therefore assume there is a special relationship between these two patterns in those few languages in which they both appear? A very preliminary statistical analysis of the co-occurrence of these patterns by Ohala and Lorentz (1977) found no disproportionate incidence of labial velar stops in languages which also have gaps in their stop inventory.

The most serious problem with such structuralist arguments, however, is that they often as not conflict with the evidence one can obtain from phonological alternations, including allophonic variation:

- 1) In spite of the double motivation mentioned above for assigning the Efik / $\widehat{kp}$ / to the labial slot (as well as an additional reason, cited by Welmers 1973, namely, that / $\widehat{kp}$ / sometimes is realized as the allophone [p]), Cook (1969) reports that a nasal assimilating to it sometimes appears as the velar nasal [ŋ].
- 2) According to Bearth (1971:18), Toura has both /k/ vs. /k<sup>w</sup>/ and /g/ vs. /g<sup>w</sup>/ contrasts, which, following the logic presented above, would force us to characterize / $\widehat{kp}$ / and / $\widehat{gb}$ / as labials. Nevertheless, these latter two sounds can be realized as [ $\text{ʔ}\widehat{kp}$ ] and [ $\text{ʔ}\widehat{gb}$ ], respectively, before nasal vowels.

Maybe one could still salvage the practice of looking only to system-internal relations in phonological analysis by abandoning the 'fill-the-gap' criteria and relying more heavily on how segments pattern in phonological rules. Unfortunately this escape route is not open either because labial velars can pattern in seemingly inconsistent ways in phonological rules.

- 3) The Yoruba labial velar glide /w/ (along with the labial velar stops / $\widehat{kp}$ / and / $\widehat{gb}$ /) patterns with the labials /b, f, m/ in that it causes the merger of following /ä/ with /ɔ̃/; nevertheless, the nasal assimilating to /w/ shows up as the velar [ŋ] (Ward 1952).
- 4) In Kuwaa (Belleh), word initial /w/ is occasionally realized as [ŋ<sup>w</sup>], i.e. a labialized velar nasal, but may become labial [v] before unrounded vowels (Thompson 1976).
- 5) In Tenango Otomi /h/ becomes labial fricative [ɸ] before /w/ but /n/ assimilating to /w/ appears as [ŋ] (Blight and Pike 1976).

Additional such cases are not difficult to find (Ohala and Lorentz 1977).

The seeming confusion of these patterns is cleared up when system-external evidence is obtained, viz., data on phonological universals and the physical phonetic causes of the universals. I offer the following four statements of universal tendencies to account for the observed data:

A. When affecting the quality of adjacent vowels, labial velars behave primarily as labials. (Specifically, they cause vowels to shift in the general direction of [u].)

In addition to the evidence in 3, above, there is that from Tigre where, due to assimilatory action, certain short vowels are more back in the environment of labials, especially /w/ (Palmer 1962).

The phonetic basis of this pattern is the fact that labial velars achieve very low 1st and 2nd formant frequencies -- even lower than those of plain labials in most cases (Ladefoged 1964, Lehiste 1964) -- and thus are acoustically unlike sounds at any other than the labial place of articulation. This fact is itself capable of being explained by reference to acoustic phonetic theory (see Ohala and Lorentz).

B. When assimilating to adjacent vowels, it is the labial velar's labial place of articulation that remains unchanged; the place of the lingual constriction may shift or disappear under the influence of the vowel's lingual configuration.

Besides the evidence in 4, above, there is in addition the pattern from Dagbani in which the phonemes /k̄p̄, ɡ̄b̄, ŋ̄m̄/ have the allophones [f̄p̄, d̄b̄, n̄m̄], respectively, before front vowels and the palatal glide /j/ (Wilson and Bendor-Samuel 1969).

There is no mystery about the causes of this tendency. Of the two constrictions of labial velars, only the lingual constriction is free to (partially) assimilate its place of articulation to that of adjacent vowels. The shift of the lingual constriction in such a case is exactly comparable to its shift in other velar consonants, e.g., [k, g, ŋ, x], whose lingual constriction -- as is well known -- is also influenced by neighboring vowels. The labial constriction, for obvious anatomical reasons, is not likely to shift its place of articulation via assimilation to that of the lingual constriction of adjacent segments.

C. When becoming a fricative or determining the place of articulation of adjacent fricatives by assimilation, [w] shows itself primarily as a labial.

In addition to the evidence in 5 (and possibly in 1) above, there are supporting statements such as the following by Heffner (1964: 160):

The fricative noises produced by the articulation of [French] [w] are slight, but such as they are, they come rather from the labial than from the velar constriction. Assuming that there are both labial and velar sources of fricative noise during these sounds, there are a number of possible phonetic reasons why the labial noise source should predominate. The most important is probably the fact that the configuration of the vocal tract anterior to the velar noise source (the airspace and the small labial constriction) constitute a low-pass filter that effectively attenuates the predominantly high frequency noise produced at the back constriction. The noise source at the labial constriction, of course, suffers no high frequency attenuation.

D. When becoming a nasal or determining the place of articulation of adjacent nasals by assimilation, labial velars behave primarily like velars.

Alongside the evidence in 1 through 5, above, there are many cases such as the dialectal variants for the word for "child" in two Melanesian languages: in Sa'a it is /mwela/ (which is more representative of the original form) but in Kwara 'Ae it is /ŋela/ (Ivens 1931).

The explanation for this pattern requires reference to the vocal tract configurations for the nasal consonants [m, n, ŋ] and [w̄] (to pick a common labial velar), which are represented schematically in Figure 1.

Essential to the acoustic characteristics of nasals are the pharyngeal-nasal airway and the oral cavity branching off from it. 'Oral cavity' here refers to that air space extending from the pharynx to the point of constriction. The oral configuration

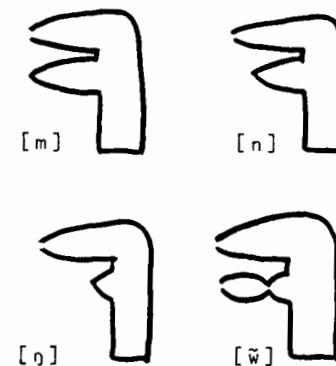


Figure 1.

anterior to the point of the rearmost constriction has no effect. It can be seen, therefore, that the acoustically relevant configuration of [w̃] is essentially similar to that of [ŋ].

It would seem from these data that the behavior of speech sounds is better understood by reference to system-external factors than system-internal factors. These are not isolated examples. A more appropriate analogy to offer as an image of language would be the game of football (American-style football). At any given time during a football game when the ball is in play, it is still the case, as in chess, that there is "significance" to the game in the special arrangement of the players, e.g. it is advantageous to the side possessing the ball to have an eligible receiver downfield. However, of more importance to the outcome of the game is the inherent ability of the individual players. It may not matter in chess whether one substitutes an ivory chess piece for a wooden one, but does matter in football if one substitutes a 50 kg tackle for one weighing 100 kg.

#### Conclusion

Observations of universal phonological tendencies -- for example, those found for labial velars, as in the present paper -- force us to the conclusion that the inherent physical constitution of speech sounds, i.e., how they are made and how they sound, have as much or more importance than system-internal relations, in determining the behavior of speech sounds. The emphasis most schools of phonology put on the study of system-internal factors is therefore a mistake.

#### Acknowledgment

The research was supported in part by the National Science Foundation.

#### References

- Anderson, S.R. (1976): "On the description of multiply-articulated consonants", JPh 4, 17-27.
- Bearth, T. (1971): "L'Énoncé Toura (Côte d'Ivoire)", Summer Institute of Linguistics, Publications in Linguistics and Related Fields, No. 30, Norman: S.I.L.
- Blight, R.C. and E.V. Pike (1976): "The phonology of Tenango Otomi", IJAL 42, 51-57.
- Chomsky, N. and M. Halle (1968): The sound pattern of English, New York: Harper and Row.
- Gamkredlidze, T.V. (1975): "On the correlation of stops and fricatives in a phonological system", Lingua 35, 231-261.
- Heffner, R-M.S. (1964): General phonetics, Madison: Univ. Wisconsin Press.
- Ivens, W.G. (1931): "A grammar of the language of Kwara 'Ae, North Mala, Solomon Islands", Bulletin of the School of Oriental and African Studies 6, 679-700.
- Ladefoged, P. (1964): A phonetic study of West African languages, Cambridge Univ. Press.
- Lehiste, I. (1964): "Acoustical characteristics of selected English consonants", IJAL, Publication 34.
- Ohala, J. and J. Lorentz (1977): "The story of [w]: an exercise in the phonetic explanation for sound patterns", Berkeley Ling. Soc., Proceedings 3, 577-799. Reprinted in: Report of the Phonology Laboratory (Berkeley), 1978, 2, 133-155.
- Palmer, F.R. (1962): The morphology of the Tigre noun, London: Oxford Univ. Press.
- de Saussure, F. (1966): Course in general linguistics. '[Transl. of original 1916 French ed.]', New York: McGraw-Hill.
- Sherman, D. (1975): "Stops and fricative systems: a discussion of paradigmatic gaps and the question of language sampling", Working Papers on Language Universals (Stanford) 17, 1-31.
- Thompson, R.B. (1976): A phonology of Kuwaa (Belleh), M.A. thesis, San José State Univ.
- Ward, I.C. (1952): An introduction to the Yoruba language, Cambridge: Heffer.
- Welmers, W.E. (1973): African language structures, Berkeley: Univ. of California Press.
- Wilson, W.A.A. and J.T. Bendor-Samuel (1969): "The phonology of the nominal in Dagbani", Linguistics 52, 56-82.

## UNIVERSALS AND PHONETIC HIERARCHY

Kenneth L. Pike, Summer Institute of Linguistics  
7500 W. Camp Wisdom Road, Dallas, Texas 75211

1. The Presumed Theoretical Basis for Some Past Avoidance of Syllable and Stress Group

In the mid 50's I cited evidence (Pike 1955, 66-68, amplified in 1967, 409-23) that on the American scene--and sometimes elsewhere also--the syllable had been often ignored, or denied theoretical status, or occasionally used without theoretical justification to support statements about the distribution of phonemes. Specifically, we might add that in Bloch and Trager (1942), in the chapter on phonetics, there is no section for the syllable (although there is one page--28--on 'Syllabic consonants' in which the syllable concept is used as background to the analysis). Similarly in the section on 'Semivowels' (22) syllabics are related to sonority, and syllables to syllabic sounds, with vowels treated as sometimes--but not always--syllabic. Later, in the chapter on phonemics, in the subsection on 'Vowels' (50) the syllable is used as a basis for discussing the distribution of simple vowels with strong stress, and related matters. But nowhere does the syllable as such get specific treatment in its own right as a basic unit of the system.

The reason: The underlying theoretical construct moved from the phoneme level to the morpheme level and on up to syntax, without the concept of syllable entering in as a level. They felt that a morpheme could be adequately described, in so far as its physical components were concerned, as made up of a sequence of phonemes. But if they had brought in the syllable as a basic unit of the system, there would have been much greater difficulty in justifying their descriptions, since oftentimes in ordinary speech a morpheme may be found which is either less than a syllable or more than a syllable, so that this leads to borders between units of the lexicon which would have been skewed with reference to those of the phonology. Thus the plural allomorph -s, is a single nonsyllabic consonant; but cups is a single syllable of two morphemes; and the morpheme ticket is a single morpheme of two syllables. Therefore, there could have been no direct mapping of (phonological) part to (morphological) whole if the syllable

had been treated as a unit in its own right.

2. A Theoretical Basis for Allowing Syllable, Stress Group, and Higher Level Phonological Units

In order to allow syllable, stress group, and even higher level units into our practical description, as units appropriate to that description, we need to have a theory of hierarchy which is multiple. Instead of a single hierarchy from phoneme to morpheme to syntactic unit, we need a hierarchy of phonology in its own right (from phoneme to syllable to stress group to phonological paragraph to phonological discourse--or something related to such a construct), and we need a hierarchy of grammatical units (from class of morphemes, to class of words, to class of clauses, class of sentences, class of paragraphs, and ultimately up to discourse classes), and in addition we need a referential hierarchy (of participants, episodes, and events as spoken about). The grammatical hierarchy (the telling order) may be distinct from the referential hierarchy (the happening--or logical--order, see Pike and Pike 1977, 363-410). Such a set of hierarchies in the theory allows us to have the syllable present in our description, and to draw upon it without apology (and without "boot-legging" it into the description).

This approach also allows us to specify openly some universals (e.g. no language is made up wholly of vowels) even though in some of them we may not find syllables composed of vowel plus following consonant. On the other hand, it does not insist that every possible level be present in every possible language. It insists, rather, that there be some hierarchical structure above the phoneme, without demanding that the syllable as such must inevitably be an emic unit. My personal suspicion would be that the syllable should be such a universal emic unit. But we have to leave room to the contrary, unless or until someone shows that the material on Bella Coola by Newman (in which the syllable is not treated as relevant) is not a satisfactory description (for preliminary discussion see Pike 1967, 420-21). Similarly, the work of Kuipers on Kabardian would have to be shown as better re-analyzed from a syllabic point of view (possibly by showing that he, like Bloch and Trager, relied on syllable without making adequate place for it in his theoretical system, for references see Pike 1967, 423).

The hierarchical approach also opens the door to the handling



of phonological markers of units much larger than a sentence (for example, the phonological paragraph). And in between the stress group and the phonological paragraph there may be emic sequences of stress groups (sequences of intonation contours) which have some overriding rising or falling general drift (or "tangent") within clause or sentence (see Bolinger 1970). And, above this, one may expect to find phonological units which signal the audience that a speaker is getting under way, or is finishing, or is changing focus. It should also be noted that there is strong evidence (overwhelmingly persuasive to me) that the kind of dynamic crescendo (or decrescendo) pattern of stress groups may in some languages be sharply contrastive within the styles of a single system. A greeting style, or a chanting style, or narrative pattern may, for example, affect these shapes; see Pike 1957, for example, for abdominal pulse types in inland Peru. A mark for juncture, plus a stress mark, is far from adequate to represent these contrasts; there must be both contrastive peaks and contrastive slopes leading down toward an end point (not just a stress mark followed by a final fade into some kind of "juncture").

### 3. Pairing in the Phonological and Grammatical Hierarchies

But the phonological hierarchy is not as simple as it sounds. There is no one direct sequence from phoneme to phonological discourse which meets some of the requirements for describing certain kinds of data which have an impact on us. Specifically, one of the most interesting developments--from my point of view--is that of Tench (1976). Tench was going beyond preliminary work on paired levels of the grammatical hierarchy (see now Pike and Pike 1977, 21-28) in which there was a sharp difference between units which are isolatable in the sense that (like an independent clause or an independent sentence) they could come at the beginning of a monologue, or at the beginning of a conversation after the greeting forms; and these would be in sharp contrast to responses to utterance, when the responses might sometimes be single words or phrases. This had led Pike and Pike to the setting up a difference between independent clause or sentence (as serving the function of serving as a proposition) versus word or phrase serving as a term. Tench showed a parallelism of these facts with the phonology, in which the syllable is the minimum independent item analogous to clause, while the rhythm group is the analogue of the

independent complex sentence. Similarly, he showed that the single phoneme (e.g. a single consonant), is analogous to a word (which is not isolatable in the same way) and that the consonant cluster would be the expanded version of that item, analogous to the phrase.

### 4. On Digital Versus Analogic Elements

More work needs to be done, also, to check out possibilities of digital versus analogic phonological structures. The digital ones (as pointed out by Martin and Pike 1975) are contrastive (either-or) units, the analogical elements have gradient (less to more) relation to the referent. My expectation would be that in every language we would find some analogic features of intonation and voice quality, in which length, loudness, rate, pause, decrescendo, crescendo (or features such as intensity, key, tenseness of vocal chords, breathiness), might be relevant in a gradient way, emphasizing the involvement of the speaker to a greater or lesser degree, or associated analogically with excitement or intensity of attitude.

But we would have to avoid assuming that such features were automatically to be found as digital in every language. For example, in Comanche (U.S.A.) no digital (contrastive, "segmentally phonemic") intonation elements have been found (Smalley 1953, 297).

The English-speaking actor on the stage, furthermore, is likely to make much greater use of the analogic types (change of key, for example), than is the ordinary person in a non-emotional setting. Yet our study of the systemic nature of contrastive quality is still in a very primitive state. It is astonishing that changes in voice quality seem to me to be empirically universal, but that a systemic handling of these materials is still only vaguely present with us. A "list" of voice qualities is far from satisfactory in handling the n-dimensional space which seems to be implicit in the possibility of simultaneous voice qualities, overlapping with pitch of various kinds, and interrupting (noncoterminal) units of the segmental phonological hierarchy from phoneme through syllable on up to phonological discourse. A vast amount of work seems to me to be awaiting us on the theoretical and empirical facets of these matters.

A final note: I am aware that there are difficulties in

finding physical correlates for perceived syllables. But I am convinced that any failures to do so in the past should not prevent us from continued search for something which is so obviously present in field work--since I cannot believe that a characteristic so universal can have no relation to some concomitant physical reality (no matter how complex the relation may prove to be).

#### References

- Bloch, Bernard, and George L. Trager (1942): Outline of Linguistic Analysis. Baltimore: Linguistic Society of America.
- Bolinger, Dwight (1970): "Relative Height", in Prosodic Feature Analysis, Pierre R. Léon, Georges Faure, and André Rigault (eds.), 109-25. (Reprinted in Intonation, Bolinger, ed., 137-53, Harmondsworth: Penguin.)
- Martin, Howard R., and Kenneth L. Pike (1975): "Analysis of the Vocal Performance of a Poem: a Classification of Intonational Features", Lg. and Style 7, 209-18.
- Pike, Kenneth L. (1957): "Abdominal Pulse Types in Some Peruvian Languages", Lg. 33, 30-35.
- \_\_\_\_ (1967): Language in Relation to a Unified Theory of the Structure of Human Behavior. Second edition. The Hague: Mouton. (First edition Vols. 1-3, 1954, 1955, 1960.)
- \_\_\_\_, and Evelyn G. Pike (1977): Grammatical Analysis. Summer Inst. of Linguistics Publ. in Ling. 53.
- Smalley, William A. (1953): "Phonemic Rhythm in Comanche", IJAL 19, 297-301.
- Tench, Paul (1976): "Double Ranks in a Phonological Hierarchy", J. of Ling. 12.1-20.

BASES FOR PHONETIC UNIVERSALS IN THE PROPERTIES OF THE SPEECH  
PRODUCTION AND PERCEPTION SYSTEMS

Kenneth N. Stevens, Massachusetts Institute of Technology,  
Cambridge, Massachusetts 02139, U.S.A.

This paper discusses how the properties of the human articulatory and perceptual systems play a role in determining certain phonetic universals. In particular, our concern is with the inventory of phonetic segments that are found in language, and the way in which these segments are organized into a set of natural classes. We shall review how the articulatory and the perceptual systems place certain constraints on the classes of sounds that are used universally in language. The classificatory features that play a role in the phonological rules in language are determined by these natural classes that are based on observation of the capabilities of the articulatory and perceptual mechanisms.

Articulatory evidence for natural classes of speech sounds.

The actualization of a given speech sound in context requires a complex sequence of articulatory activity. The articulatory structures must be maneuvered from positions or states appropriate to one sound to states corresponding to the next sound to be produced. We shall follow the traditional method, used by phoneticians for years, of specifying a phonetic segment in terms of a set of goals or target states that the articulators are to achieve or that are intended by the speaker rather than in terms of the movements between these targets. The hypothesis is that these target configurations or states, if appropriately specified for a given sound, are much less dependent on the phonetic context than are the articulatory movements or muscle contractions necessary to produce the sound in context. Thus the articulatory descriptions are static, in the sense that they describe stationary states or configurations. While the production of some sounds or sound sequences may involve movement, this movement is always from one target state to another.

How are these articulatory target states to be described and how does this description lead to a specification of natural classes of speech sounds? Examination of lateral radiographs

gives us one view of the articulatory target states in terms of the positions of the various articulatory structures that are visible on the midline. This kind of evidence has traditionally been used in phonetics to describe articulatory targets in terms of place of articulation identified along the length of the vocal tract. Another way of describing articulatory configurations examines the pattern of contact that occurs between structures such as the tongue and palate. This pattern is presumably registered in the talker's speech control system through the responses of receptors located on the surfaces of the structures (Stevens and Perkell, 1977). Still another aspect of the target state is the physical properties of the surfaces of the structures, particularly the vocal folds and the tongue. These properties have an influence on the manner in which the airflow from the lungs is controlled and on the way in which the articulatory structures are forced against one another. Which of these ways (or combinations of ways) of describing articulatory states is most salient for grouping speech sounds into natural classes is a question about which we can only speculate at present.

We consider now several lists of phonetic segments. For all of the items on a given list, some aspect of the articulation is achieving the same state, as defined in at least one of the ways listed above. We suggest, then, that these items can be candidates for forming a natural class of phonetic segments.

[m n ŋ ã õ ...] These items are all produced by creating velopharyngeal opening, usually by placing the velum in a lowered position. From the point of view of the speaker, an indication that the velum is lowered comes from several possible sources: (1) the muscles used to lower the velum have been contracted; (2) the lowered state of the velum is sensed through receptors that signal the position of the velum or its contact with other structures; (3) there is airflow through the velopharyngeal opening and possibly acoustic energy in the nasal cavity that is sensed and registered in some way.

[k g ŋ i u ...] These sounds are all produced by placing the tongue body in a raised position within the oral cavity. More specifically, the common articulatory activity for the sounds can

be described in one of two ways: (1) there is contraction of a common muscle or group of muscles to produce the raised tongue body, or (2) there is a common pattern of activity in particular groups of sensory receptors in the tongue musculature or on the dorsal surfaces of the tongue as these surfaces make contact with other structures, particularly the hard palate (Stevens and Perkell 1977).

[p t k č f θ s š á í ú ...] For this group of sounds, it is hypothesized that the common articulatory attribute is a stiffening of the surfaces of the vocal folds (Halle and Stevens, 1971). The articulatory state that characterizes each member of this class can be described either as contraction of a particular laryngeal muscle or group of muscles or as the stiffened state of the vocal fold surfaces, independently of the muscle activity used to produce that state.

[p t k č b d g ĵ m n ŋ ...] The sounds in this group are all produced by forming a complete closure of the vocal tract at some point along its length. The articulatory description for this group of segments cannot be specified in terms of the contraction of particular muscles, since different muscles are clearly involved depending on where in the vocal tract the constriction is made. Rather, it is assumed that an instruction to form a complete closure is a basic component of articulatory control which, when coupled with a further instruction indicating which articulator is to be activated, effects the proper consonantal constriction. It is possible also that the sensory consequences of forming a complete closure are registered in some unique manner independently of the location of the closure in the vocal tract.

[p b f v m ...] The segments on this list have the common articulatory attribute that they are produced with a constriction at the lips. Thus a particular set of muscles - those making a lip closure - is involved in the generation of all of these sounds. The lower lip comes in contact with either the upper lip or the upper incisors, and this gesture leads to a unique pattern of excitation of sensory units in the lower lip.

[t d n θ ó s z š ž ʎ r ...] These phonetic segments are all actualized by raising the tongue blade to make contact with some

part of the maxilla. The exact region of contact or the force of contact may vary from one sound to another in the set, but the common gesture is that of raising the tongue blade, presumably through contraction of certain intrinsic tongue muscles. There is a unique sensory consequence of this raised pattern of the tongue blade: the edges of the superior portion of the tongue come in contact with fixed surfaces of the hard palate or teeth, presumably leading to a special response of tactile receptors on these surfaces of the blade.

The six lists of segments given above are examples of a longer inventory of lists of segments that could be generated. Furthermore, there is no attempt to make each list exhaustive; additional items could be appended to the lists. These examples serve to indicate, however, that natural classes of speech sounds can be constructed through examination of the articulatory target configuration or states. In giving these examples, we have shown a certain amount of ambivalence as to how the common articulatory attributes for the items on a list should be specified. Until we know more about how motor systems operate, and, in particular, how the speech-production systems operate, the question of how best to characterize natural classes of speech sounds in terms of articulatory attributes must remain open.

#### Acoustic and psychoacoustic evidence for natural classes

Acoustic analysis of speech shows that there are groups of speech sounds that share common acoustic properties. If it is assumed that the auditory system responds in some unique way to sounds with a common acoustic property, then this unique response provides the listener with a means for organizing speech sounds into natural classes based on their acoustic properties. As examples, we shall consider several lists of speech sounds, and we shall show that for the items in any one of these lists there is a common distinctive acoustic property. The basis for these classifications is derived largely from the work of Fant (1960), Jakobson, Fant and Halle (1963), and others.

[m n ŋ] For the items on this list, there is a rather steady nasal murmur persisting for several tens of milliseconds, with an amplitude just a few dB below that of the adjacent vowel. The unique acoustic attribute of this nasal murmur is a strong

spectral peak at low frequencies and a relatively uniform distribution of weaker spectral peaks at higher frequencies, with these peaks tending to be rather broad (Fujimura, 1962).

[t d n s z ʒ ʒ̥ ʃ ʃ̥] For these consonants, the spectrum sampled at or near the consonantal release (in a consonant-vowel syllable) shows a diffuse spread of energy across the frequency range, but with greater spectral energy at high frequencies (Fant, 1960; Zue, 1976; Stevens and Blumstein, 1978).

[k g ŋ] The spectrum at the consonantal release for these sounds has a single prominent peak in the midfrequency range (Fant, 1960; Zue, 1976, Stevens and Blumstein, 1978).

[i ɪ u u] The vowels in this list all have a relatively low first formant.

[ã ü ĩ] These nasalized vowels have a spectrum in which the lowest peak, corresponding to the first formant region for a nonnasal vowel, is split or broadened to cover a wider frequency range than that for a nonnasal vowel.

[p t k ʧ b d g ʝ m n ŋ] The items in this list all show an abrupt onset of spectral energy over much of the frequency range when the consonant is released into the following vowel. The rise in amplitude in any one frequency region occurs in a time interval of just a few milliseconds. A sound with an abrupt onset has been shown to produce a distinctive response in a listener (Cutting and Rosner, 1974).

[á ú í] These vowels all have a fundamental frequency ( $F_0$ ) that is high in comparison with the average  $F_0$  for the particular speaker and the particular position of the vowel within an utterance.

[p t k ʧ f θ s ʃ] The common acoustic characteristic of the sounds in this list is the absence of low-frequency periodicity in the sound in the vicinity of the consonantal closure interval.

As in the case of the lists based on articulatory attributes, the above lists are examples of a longer inventory of lists such that the items in each list have a common acoustic property to which the auditory system is assumed to respond in a unique way. Given our present rudimentary knowledge of the response of the auditory system to complex sounds, we have only

been able to speculate on the kinds of acoustic properties that qualify for defining groups of speech sounds.

#### The classificatory features

Examination of the two sets of lists - these based on common articulatory attributes and those based on common acoustic attributes - reveals that there is much overlap in the two sets. This overlap is not surprising, since on the basis of acoustical theory it is not unexpected that sounds produced with common aspects of the articulatory configuration should also have similar acoustic characteristics.

Another way to organize speech sounds into natural classes is to examine the phonological rules of language, and to observe the various groups of segments that are operated on by these rules or that determine the environments in which the rules operate. The grouping of segments according to this criterion leads to a description of segments in terms of bundles of classificatory or distinctive features. These classificatory features also show a great deal of overlap with the groupings based on articulatory and acoustical considerations.

We would like to propose a rather simple condition on the definition of a classificatory feature: a set of speech sounds shares the same classificatory feature if the sounds share a common articulatory attribute and a common acoustic or perceptual attribute. That is, the sounds in a given class should give rise to response patterns that have a common property in the auditory system of the listener and the speaker, and, in addition, the production of the sounds should have common attributes in the speech-generating mechanism of the speaker, such as common patterns of orosensory response.

A consequence of this definition is that vowels and consonants will tend not to share the same features. Thus, for example, nasal vowels and nasal consonants would not have the same feature, although it might be desirable to mark in some manner the fact that they share an articulatory property. The strong definition of a classificatory feature would not capture in terms of feature specifications the fact, for example, that vowels preceding nasal consonants tend to be nasalized (or in fact that nasalization of the vowel often is accompanied by

elimination of the consonant), or the fact that the pitch of vowels following voiceless consonants tends to be raised. These kinds of modifications are, in a sense, simply mechanical consequences relating to the coarticulation that is a nature consequence of the juxtaposition of two segments.

The classificatory features defined in the way we have proposed would, however, specify major classes of segments that play a role in the phonological rules of language. These features would owe their existence, so to speak, both to the property-generating characteristics of the speech production system and to the property-detecting characteristics of the speech perception system.

#### References

- Cutting, J. and B. Rosner (1974): "Categories and boundaries in speech and music", Perc. Psych. 16, 564-570.
- Fant, G. (1960): Acoustic theory of speech production, The Hague: Mouton.
- Fujimura, O. (1962): "Analysis of nasal consonants", JASA 34, 1865-1875.
- Halle, M. and K.N. Stevens (1971): "A note on laryngeal features", Research Laboratory of Electronics Quarterly Progress Report No. 101, M.I.T., Cambridge, Massachusetts, 198-213.
- Jakobson, R., G. Fant, and M. Halle (1963): Preliminaries to speech analysis, Cambridge, Massachusetts: M.I.T. Press.
- Stevens K.N. and S.E. Blumstein (1978): "Invariant cues for place of articulation in stop consonants", JASA 64, 1358-1368.
- Stevens, K.N. and J.S. Perkell (1977): "Speech physiology and phonetic features", in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.), 323-341, Tokyo: University of Tokyo Press.
- Zue, V.W. (1976): Acoustic characteristics of stop consonants: A controlled study, Ph.D. thesis (unpublished), M.I.T., Cambridge, Massachusetts.

## THE PSYCHOLOGICAL REALITY OF PHONOLOGICAL DESCRIPTIONS

## Summary of Moderator's Introduction

Victoria A. Fromkin, University of California, Los Angeles,  
California 90024, USA

A phonological description of a language will be a 'true' description to the extent that it is 'psychologically real'. A theory of phonology will be a 'true' theory to the extent that it permits the construction of psychologically real grammars. These assumptions are required of an empirically based phonological theory. What we seek then is evidence that will help decide whether a particular description is 'psychologically real'. There are no a priori principles which can be depended on. We do not know in advance whether, for example, the human mind can or does relate two levels of phonological representation--phonemic and phonetic--by ordered rules, nor do we know the extent to which the immature child's brain can draw highly abstract generalizations from a limited set of input stimuli. In fact, we have not progressed too far since 1887 when Fournie observed that "Speech is the only window through which the physiologist can view the cerebral life". Psychologists, neurologists, and linguists depend, to a great extent, on linguistic facts to determine the capabilities of the human mind. We have not found any direct ways, as yet, to observe what is "'in people's heads' (and) since we cannot look into people's heads directly we can only hypothesize what goes on there on the basis of indirect evidence" (Chafe, 1970). Even when we do look into people's heads directly, we cannot find in the physical brain matter, in the  $10^8$  neurons, or even in the neural organization of the cortex, the information we seek regarding the nature of the internalized grammars, the information which will tell us whether our theory, or which theory, of phonology is psychologically real or 'true'.

This symposium is concerned with the kinds of evidence which will help decide this question. While we all seem to agree on our aims (at least to the extent that we seek 'psychological real grammars') we are not necessarily in agreement as to what counts as evidence, how to weigh different kinds of evidence, or even what is meant by 'psychological reality'.

Cutler suggests a division between the proponents of a

'strong sense' as opposed to a 'weak sense' of psychological reality. The first group considers levels (e.g. phonemic representations) and processes (e.g. P-rules) to be psychologically real if a processing model includes stages isomorphic to levels and mental operations corresponding to the processes or rules. Linell also refers to this division. Cutler's paper presents speech error data to show that lexical stress and word formation rules are psychologically real in the weak sense, but not in the 'operational' or 'strong' sense. Linell also suggests that "rules must not be equated with behavioral processes... (since) conventional phonological rules state nothing but regular correspondences between idealized representations of the same or related pronunciations." In the fuller version of my paper I will discuss some evidence from speech errors which suggests that at least some rules and some levels are real in the strong sense of the term, but that this should not be a criterion for a theory of phonology.

Derwing's paper seems to support the 'strong' view. For example, he questions "what psychological sense can possibly be made... of a notion of 'rule ordering' which has no relation to real time" and further proposes that "if grammars relate in any way to psychological events or states (my emph.) then we need to interpret grammars psychologically." Grammars can, however, 'relate' to events or states without being identical or even isomorphic to them. And one can conceive of ordered relations, hierarchical for example, in a non-behavioral way and on a non-real-time basis. The alphabet may be represented in memory ordered from A to Z even for a brain damaged patient who cannot retrieve the letters in that order in real time. Cognitive psychologists concerned with lexical storage are providing evidence for intricately ordered classification systems based on ordered basic and primary levels of categorization in the levels of abstraction in a taxonomy (Rosch, 1978). Derwing also discusses aspects of the question which relate to the philosophy of science (as do Linell and Skousen), some points of which I will further discuss. But it is clear that whether a theory or a grammar is psychologically real must depend on empirical evidence rather than one's philosophical biases.

Bondarko's paper is neutral as to some of the controversies discussed in the other papers, positing three psychologically

real levels of phonology--production and perception of speech sounds, the phonemic level, and the level of word formation rules--as evidenced by perception experiments.

Campbell, Dressler, Gussman, and Skousen, are concerned with the importance of internal versus external evidence in the testing of linguistic hypotheses and the evaluation of theories. Internal evidence refers to facts drawn from the overall grammar, significant generalizations, simplicity factors, distributional criteria, morphemic alternations etc. External evidence refers to acquisition data, language disturbance, borrowing, orthography, speech and spelling errors, metrics, casual speech, language games, historical change, perception and production experiments etc. (Cf. Zwicky, 1975). Campbell and Skousen, and to a certain extent, Dressler, place major emphasis on external evidence. Campbell is very convincing in his demonstration of how language games in Finnish and Kekchi, for example, strongly support the reality of a vowel harmony rule and a vowel-epenthesis rule, respectively. He provides similar evidence in support of morpheme structure conditions as opposed to syllable structure rules. Skousen uses similar arguments. But Dressler shows that external evidence can be contradictory and Gussman provides some detailed illustrations supporting this. Interestingly, where Skousen posits external evidence from tongue slips to show the correctness of analyzing the affricates in English as non-sequential units, /č/ and /ǰ/, Gussman provides other external evidence, i.e. low level phonetic rules, which argue for the sequential analysis. Gussman points to the Fromkin (1971) data cited by Skousen to illustrate this contradiction. He also ties in the question of 'abstractness' with 'psychological reality' and correctly, I believe, shows that the question should not be how abstract is an analysis, but is it right or wrong. An important question to be discussed in the symposium, then, is what to do when different kinds of evidence are contradictory. It is also important for us to clarify how both internal and external evidence are to be used. If we find in Kekchi, for example, that an experiment on loan words supports morpheme structure conditions is this to be used only for the grammar of Kekchi or as evidence for the meta-theory of phonology? If speech error data argue for a rather abstract representation in some language, is this evidence that one can provide such abstract



representations in all languages? In other words, are we looking for evidence as to constraints on a general theory of phonology or for evidence concerning a grammar of a particular language?

Given the extent to which individual grammars may vary across speakers of one language, should we not seek constraints on the general theory which will permit us to construct the optimal, 'psychologically real' grammar for a language? The papers already cited reveal the problems we face. Data alone, and multiple-kinds of evidence alone will not provide all the answers. We need universal principles and a theoretical framework which in a principled fashion will help us constrain phonological descriptions to psychologically real ones. Skousen presents such a principle-- a principle of maximizing acoustic differences. Hale's paper is primarily concerned with just such questions and posits a 'principle of recoverability', with supporting evidence from Papago and Maori. What we need is more principles, supported by clear empirical evidence. For we can probably all agree that "However difficult it may be to find relevant evidence for or against a proposed theory, there can be no doubt whatsoever about the empirical nature of the problem" (Chomsky and Halle, 1968).

#### References

- Chafe, W. (1970): Meaning and the Structure of Language, University of Chicago Press, Chicago.
- Chomsky, N. and M. Halle (1968): The Sound Pattern of English, New York: Harper and Row.
- Fromkin, V.A. (1971): "The non-anomalous nature of anomalous utterances", Language 47, 27-52.
- Rosch, E. (1978): "Principles of categorization", in Cognition and Categorization, Eleanor Rosch and Barbara B. Lloyd (eds.), 27-48, Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Zwicky, A. (1975): "The strategy of generative phonology", in Phonologica 1972, Dressler and Mareš (eds.), 151-168.

## ON THE PHONOLOGICAL OPERATIONS ENSURING SPEECH COMMUNICATION

L.V. Bondarko, Department of Phonetics, University of Leningrad, USSR

Conveying information through articulate speech presupposes the ability of the native speaker to analyse quickly and effectively heterogeneous sounds. This ability is developed by man because sound differences are used for discriminating meaningful units, i.e. words. Taking this function of speech sounds into consideration, we can understand why the native speaker does very well in the process of perception in spite of a number of variations of sound properties. From the linguistic point of view it can be assumed that there exist a number of levels ensuring optimum processing of sound signals. The first one consists in the ability of man to generate and perceive articulate sounds. Though this ability is universal by itself, it cannot be observed directly because it is realized on the basis of a certain concrete language. However, some of the phonetic universals (Greenberg, 1966) deduced on the basis of comparing various languages, can be also related to the peculiar properties of man's verbal behaviour. The second level is concerned with the system of phonemes in a given language. The native speaker disposes of the information of the system of phonemes which he acquires in the process of learning his native language. The main points of this information are as follows: the inventory of the phonemes in the language, the ways the distinctive features of the phonemes are realized, the rules of usage which include the probability of the occurrence of phonemes within the minimal meaningful unit - within a word.<sup>1</sup>

The third level deals with the information of the rules about possible sound combinations in shaping the words. One can assume that the perception of the word is the recognition of its phonemic composition. Evidently a clear-cut differentiation of all the three levels is impossible, because practically they overlap to a great extent. But one may hope that the systematic research on the process of perception will enable the scientists to describe these levels in a more detailed way.

---

(1) It is possible that in a number of cases a morpheme may be treated as this minimal unit. This may take place in languages where phonemic alternations are regular and are governed by the existing rules, Russian being an example.

Let us consider some facts dealing with each of these levels which testify to the reality of the language consciousness of the speakers. The opposition of consonants with regard to "absence - presence of voice" is one of the most widespread (Zhivov, 1976). In fact, it can be connected not only with the function of the vocal cords alone, but also with properties like tenseness - laxness, delay in the onset of voice after the opening of the occlusion, the duration of the preceding vowel, and so on. One may assume that "absence - presence of voice" can be treated as a universal feature. For the native speaker of the Russian language, where the correlation "presence versus absence of voice" is one of the characteristic features, each consonant he hears must be described either as a voiceless or as a voiced one. But the consonants /c/, /č/, /x/ do not have voiced correlates, i.e., the opposition of voiceless consonants to voiced ones is not possible for them in the positions before vowels and consonants. Compare [tu'goj] - [du'goj], [sɪpɪtʃ] - [gɪbɪtʃ] and [tsex], [tʃaj], [xot], and so on. However, in accordance with the rules of alternations which are known to be regular in the Russian language, in the combination of words ending in the consonants /c/, /č/, /x/ ([ts, tʃ, x]) with words in which initial consonants are voiced obstruents, there appear voiced allophones of these voiceless consonants: [kan'nedz zɪ'mɪ], [zedz drɐv'va], [moɣ ga'ʃit], phonologically: /kan'éc z'i'mɪ/, /žeč drɐv'vá/, /moɣ gar'ít/.

The voiced character of these phonologically voiceless consonants can be treated in various ways from the linguistic point of view. We are especially interested in how the voiced character is treated by the Russian native speaker who is expected to discriminate between voiceless and voiced consonants and who does not have at his disposal the voiced correlates of phonemes which possess the same properties as /c/, /č/, /x/.

Russian subjects when presented with the consonants from phrases of the type /kan'éc z'i'mɪ/, /žeč drɐv'vá/, /moɣ gar'ít/, cut out from the magnetic tape, recognized these consonants as voiced ones; other properties of the consonants could be perceived incorrectly in this case. If the phonetic context is enlarged and the subjects are presented with combinations - 1: including the following consonant (CC), 2: including also the preceding vowel (VCC), 3: including the vowel in the succeeding syllable as well, - the

recognition of the consonants under consideration as voiced ones occurs less frequently, though in these cases the consonants /c/, /č/ and /x/ are not interpreted 100% correctly.

Figure 1 presents data on how separate properties of the consonants /c/, /č/ and /x/ are perceived if they are presented in various contexts, such as C, CC, VCC and VCCV. The influence of the phonetic features proper increases with the narrowing of the phonetic context, although even if there is a complete phonetic context - the following consonant bringing about voicing, or vowels, ensuring as a rule good recognition of the neighbouring consonant - this is not sufficient for the recognition of such phonemes as /c/, /č/ or /x/. The sounds may be perceived as /c/, /č/ or /x/ only if the native speaker hears the whole phrase, i.e. if he makes use of both the phonetic and the semantic contexts (Bondarko, 1975). This means that the predominant influence of the first, universally phonetic level is removed only if both the second level including rules of alternations, and the third level concerned with the

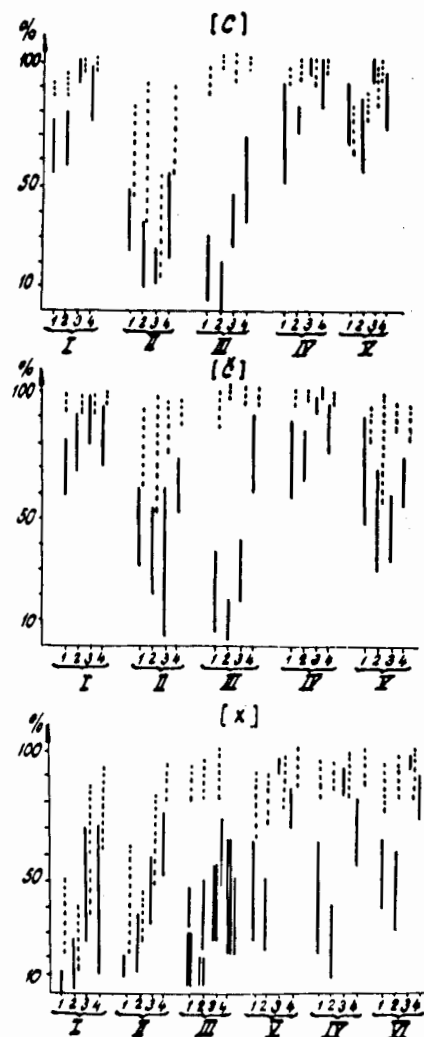


Figure 1

The perception of the properties of the voiced (—) and voiceless (---) allophones of the consonants /c/, /č/ and /x/. The phonetic context: (1) -C, (2) -CC, (3) -CCV, (4) -VCCV. Properties: I the active organ of speech II the manner of production III absence - presence of voice IV noise - sonorous character V hardness - softness VI vowel-consonant character

analysis of the phonemic composition of words can be made use of.

The second level of analysing speech, as has already been mentioned, includes information about the inventory of phonemes in the given language, the ways in which the distinctive features are realized, and the rules of usage. It is this level that ensures the transition from the phonetic variations of real sounds to economic phonological interpretations. Let us consider this level of perception using the examples concerning the perception of vowels by Russian native speakers.

It is known that the system of vowels in the Russian language is comparatively poor. There are three degrees of height and two series. Vowels of the back series (with the exception of the lowest vowel /a/ are necessarily rounded, whereas this connection does not exist in the case of the front vowels. The six vowels /a/, /o/, /u/, /e/, /i/, /i<sup>2</sup>/ are realized differently in the stream of speech, depending on their stressed or unstressed character, the quality of the neighbouring consonants, and so on.

As was shown in an experiment (Bondarko et al., 1966), the i-like transition, appearing in the vowel under the influence of the soft neighbouring consonant, serves as a useful indication which enables a person to differentiate hard and soft consonants. The i-like transition (phonetically pushing forward the vowel into the front zone) is perceived by all Russian native speakers as a cue of the consonant. Nevertheless, the phonetic property itself is realized in the vowel, and Russian native speakers discriminate a greater number of vowels than could have been expected on the basis of the inventory of vowel phonemes in the language.

We can assume that it is this peculiarity in the realization of the feature of softness in consonants that enables Russian speakers to describe vowels of the type [y], [ø], [œ] at a universal, phonetic level. These are integrated in the inventory of vowels in the same way as is done by speakers of those languages in which these vowels represent phonemes (Slepokurova, 1971). Things are different in the situation where vowels adjacent to nasal sounds are presented. In this phonetic position, Russian

vowels are considerably nasalized and it could be expected that Russian speakers would use such changes in vowels by analogy with those that are observed in the position with the neighbouring soft consonants. But in reality, the results are quite different.

In a special investigation (Belyakova, 1977) dealing with the perception of nasal vowels of the French language and nasalized Russian vowels by Russian and French subjects, it was shown that French people recognize nasal vowels of their own language much better than Russians do theirs, but that they are less sensitive in the perception of Russian nasalized vowels. They perceive Russian nasalized vowels as non-nasalized. A comparatively low degree of the recognition of the Russian vowels a and e by French listeners can be accounted for not by the influence of nasalisation but by the influence of the neighbouring soft consonant, which leads to the perception of this vowel as more front and less open, i.e. a as e, e as i. It is typical of Russians to make a lot of mistakes in the recognition of the nasalized vowels (Fig. 2).

Finally, it is on the third level, dealing with the rules of the formation of the sound shape of the word, that a phonological interpretation of sounds is given, which has no unique phonetic correlate. For example, the recognition of the unstressed vowel

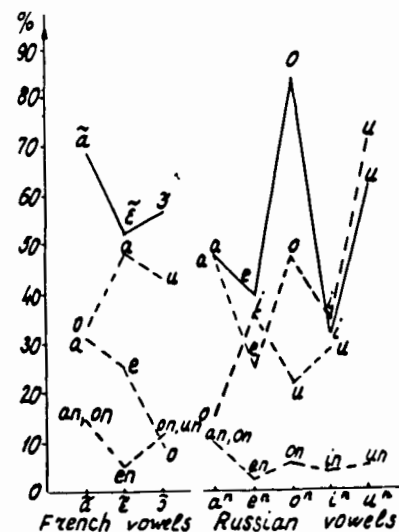


Figure 2

The perception of French nasal and Russian nasalized vowels. French listeners ——— Russian listeners ----- For all the subjects, various identifications of the vowels are shown: as the corresponding nasal vowel, as non-nasal but having different quality, and as a combination of a non-nasal vowel with a nasal consonant. Such identification is indicated in the figure by: an, en, and so on, even in those cases where the subjects wrote down the sounds an, am, etc.

(2) We do not consider here the question of the phonemic relevancy of the opposition of /i/ - /i<sup>2</sup>/, because it is widely discussed from the linguistic point of view, and, practically, because in the linguistic analysis it is not treated from the point of view of the phonology of the native speaker, for whom these are different vowels, and not on the lowest level alone.

in the words [sʌ'rok], [dʌ'ma] and so on, as /a/ is connected with the rules of reduction in the Russian word; the recognition of the voiced affricate as a voiceless one in the phrase "otec bolen" ([ʌ'tɛdz 'boʎɪn]) is connected with the rules of alternating voiceless and voiced consonants.

The recognition of morphologically loaded sounds or sound combinations represents a special case, particularly for such a language as Russian (Bondarko et al., 1966). In these cases the phonetic information about the sound is often insufficient, although the use of the rules of alternation and the use of semantic redundancy of the context enable the subject to correctly interpret the phonemic composition of the word (compare the realization of the phoneme /s/ in the combination "brosj ŝumetj" ([broʂ ʃu'mɛtʃ]) with a considerable assimilation of /s/ to the following /ʂ/ and the realization of the phoneme /a/ in posttonic inflections after the soft consonant "njanja" ([ 'nʲaŋʲɪ]), and so on.

All this proves that in oral communication, a person performs rather complicated operations the total of which can be called the phonology of the native speaker.

The reality of other purely linguistic phonological descriptions is proven by the extent to which this description is in accordance with these operations. The description of the phonology of the native speaker, based upon the description of different levels determining his verbal behaviour and upon the comparison with the linguistic phonology set up in linguistic descriptions, can be considered the main task in the experimental phonetic investigations dealing with speech perception.

#### References

- Belyakova, G.A. (1977): "The nasalization of vowels and its perception (on the basis of French and Russian languages)", Vestnik L.G.U., No. 8.
- Bondarko, L.V. (1969): "The syllable structure of speech and distinctive features of phonemes", Phonetica 20.
- Bondarko, L.V. (1975): "The phonemic description of the utterance - the condition and the result of understanding context", The minutes of the Fifth All-Union Congress of Psycholinguistics and the Theory of Communication, Moscow.
- Bondarko, L.V., L.A. Verbitskaya, L.R. Zinder and L.P. Pavlova (1966): "The sound units that can be distinguished in Russian speech", in The Mechanism of Speech-Production and Speech-Perception of Complex Sounds, Nauka, M.L.

Bondarko, L.V. and L.A. Verbitskaya (1975): "Factors underlying phonemic interpretation of phonetically non-defined sounds", in Auditory Analysis and Perception of Speech, London: Academic Press.

Greenberg, J. (1966): "Synchronic and diachronic universals in phonology", Language 42, No. 2.

Slepokurova, N.A. (1971): "On the position of the phonemic boundary between synthetic vowels", in The Analysis of Speech Signals by Man, Leningrad.

Zhivov, V.M. (1976): "The universals of syntagmatic functioning of the feature of voiceness", The Institute of the Russian Language of the Academy of Sciences of the U.S.S.R., A Study Group dealing with experimental and applied linguistics, preliminary publications, Issue 89, Moscow.

THE QUEST FOR PSYCHOLOGICAL REALITY: EXTERNAL EVIDENCE IN PHONOLOGY

Lyle Campbell, State University of New York at Albany, Albany, New York, U.S.A.

The principal goal of linguistic description is to account for a language in a way which reflects the competence of its speakers. This goal of achieving psychologically real descriptions (grammars) is reasonable. However, generative phonologists too frequently assumed that child language learners were somehow constrained to acquire the simplest possible grammar, and because the notation was designed to convert true generalizations into considerations of length, the simplest grammar the linguist could write was taken to be the psychologically real grammar. For this reason, arguments based on formal simplicity alone were considered sufficient to resolve many issues, e.g. abstractness, rule ordering, etc. However, formal simplicity (internal evidence) is not sufficient. Questions about psychological reality and the learnability of rules cannot be answered through considerations of surface patterns, distribution of allomorphs, combinatory properties of phonological elements within a linguistic system, and the like. The real question is, how can the linguist be certain that the rules he postulates to account for phonological patterns he perceives in his data correspond to the rules that speakers of the language establish? There are at present no successful formal criteria for determining from a given set of data what the speaker's rules may be.

A serious search for answers to this question must involve "external evidence", evidence not confined to surface-pattern regularities, but evidence which shows speakers behaving linguistically in ways where they must call upon their knowledge of the rules and underlying forms of their language in overt and revealing ways. The goal of this paper is to argue that external evidence should be given a stronger role in phonological investigation and to illustrate its potential.

Sources of external evidence that have been used with some success are: Metrics and verse (Kiparsky 1968, 1972; Zeps 1963, 1973), word games (Sherzer 1970; Campbell 1974, 1977) borrowing (Campbell 1974, 1976), experiments (Ohala 1974), speech errors

(Fromkin 1971, 1973), construction of orthographies, language change, etc. Here I will attempt to show the relevance of external evidence for validating aspects of individual grammars and for refining linguistic theory.

External evidence can demonstrate the psychological reality of certain rules. I will consider two examples, both involving word games (secret languages). The first is vowel harmony in Finnish. Since Finnish vowel harmony has many exceptions and complications, some have suspected that it may not be a psychologically real rule of Finnish grammar. In *kontti kieli* (or *kontin kieli*) "knapsack language", one of several Finnish word-game secret languages, the first consonant(s) and vowel of a word are replaced by ko (of kontti), and the material for which ko is substituted is placed before ntti (of kontti). Thus veitsi "knife" becomes koitsi ventti, susi "wolf" is kosi suntti. In this language game, vowel harmony adjusts the remaining vowels of the word to agree with ko (the harmonic series are back a, o, u, front ä [æ], ö [ø], y [ü], and neutral i, e):

pysähtyköön "let him stop" becomes kosähtukoon pyntti  
 kylpylössä "in the baths" becomes kopuloissa kyntti  
 hänkö "him?" becomes konko häntti

If vowel harmony were not a psychologically real rule of Finnish, speakers would not be able to adjust the vowels productively to agree with the back vowel when ko is substituted. (For full arguments, see Campbell 1977, in press.)

The second case is from Kekchi (a Mayan language of Guatemala). In *Jerigonza*, the Kekchi word game, one places p after each vowel followed by a copy of that vowel; for example q'eqi<sup>v</sup>?, the name of the language, is q'epeqipi<sup>v</sup>? in *jerigonza*. This game shows several Kekchi rules to be psychologically real, for example the rule of vowel-epenthesis before voiced labials ( $\emptyset \rightarrow V_1/V_1C \{b^m\}$ ) (examples: kwiq'ib'a:nk /wiq'-b'ank/ "to break it", k'oxob'a:nk /k'ox-b'ank/ "to seat something"). In normal speech, these forms never occur without the epenthetic vowel, but one may speak *jerigonza* optionally leaving out the epenthetic vowel: kwipiq'b'apa:nk or kwipiq'ipib'apa:nk, and k'opoxb'apa:nk or k'opoxopob'apa:nk. The rule of vowel epenthesis must be psychologically real; speakers must know the rule because they take it

into account in producing jerigonza forms -- they never leave out the wrong vowel, only the vowel which results from the rule of epenthesis.

These word games provide evidence for the reality of several other rules in these two languages, as well. Here, the external evidence helps resolve issues concerning the correct description of the individual grammars. External evidence has important implications for theoretical issues, also. I will present just one example, also from Kekchi.

Bilingual informants in Spanish and Kekchi were presented a list of loan words, some from Spanish into Kekchi and some from Kekchi into Spanish, and asked to judge whether the forms were borrowed, and if so, which they thought was the original language. Judgements were based on several parameters (cultural, semantic, and phonological). These parameters were determined by asking the informants why they thought particular loans to be Spanish or Kekchi in origin. Reasons volunteered by these informants involved, among other things, native views of morpheme structure in the two languages. For example, informants said pio:c̣ "pickaxe" (from Spanish piocha) and vilte:p "small chile" (from Spanish chiltepe), and similar forms, were from Spanish because Kekchi does not have those kinds of sounds together (vowel clusters in the first case, consonant clusters in the other). In actual fact, Kekchi does have vowel-vowel and consonant-consonant clusters, but only across morpheme boundaries (e.g. ke-ok' "get cold", ke- "cold" plus a verbal suffix), but never within a morpheme. This shows that these morpheme structure conditions of Kekchi are psychologically real, since speakers actively called upon them in making judgements about the origin of lexical items. To be sure, this evidence helps validate aspects of Kekchi grammar, namely its morpheme structure conditions. (For details, see Campbell 1974, 1976).

Perhaps more importantly, however, this external evidence shows that morpheme structure conditions are real, and cannot be accounted for merely by syllable structure rules as proposed by "Natural Generative Phonologists" (Hooper 1975). Thus external evidence provides the means for testing theoretical claims. External evidence has been shown to have important implications for several issues in linguistic theory, e.g. the controversy over

extrinsic ordering of rules, abstractness, morpheme structure conditions, etc. (See Campbell 1974, 1976, 1977; Kiparsky 1974.)

To conclude, psychological reality can be investigated empirically but it takes more than ransacking a body of data for the internal patterns and processes a linguist might find. It requires that evidence outside these internal patterns be sought which shows speakers using the rules of their language productively. As more and more cases of external evidence are considered, important issues in phonological theory may be resolved, and the answers to important questions found, questions such as: how different from the surface may underlying forms be and still be learned by speakers?; how many forms must illustrate a rule before speakers learn the rule rather than the variant forms piecemeal?; how do exceptions, non-productivity, non-phonetic conditioning factors, "opacity", and the like affect the learnability of rules?, etc. To answer these and related questions, we need sufficient external evidence, and until we answer them, phonological theory will be found wanting.

#### References

- Campbell, L. (1974): "Theoretical implications of Kekchi phonology", *IJAL* 40, 59-63.
- \_\_\_\_\_. (1976): "Linguistic acculturation: a cognitive view", In: *Studies in Mayan linguistics 1*, M. McClaran (ed.), American Indian Culture Center, UCLA.
- \_\_\_\_\_. (1977): "Generative phonology vs. Finnish phonology: retrospect and prospect", In: *Papers from the Transatlantic Finnish Conference*, R. Harms and F. Karttunen (eds.), 21-58, Texas Linguistics Forum, 5. Austin.
- Fromkin, V. (1971): "The non-anomalous nature of anomalous utterances", *Lg.* 47, 27-52.
- \_\_\_\_\_. (1973): *Speech errors as linguistic evidence*, (Janua Linguarum, Series Maior, 77), The Hague: Mouton.
- Hooper, J. (1975): "The archi-segment in natural generative phonology", *Lg.* 51, 536-60.
- Kiparsky, P. (1968): "Metrics and morphophonemics in the Kalevala", In: *Studies presented to Professor Roman Jakobson by his students*, C. Gribble (ed.), 137-48, Cambridge, Mass.: Slavica.
- \_\_\_\_\_. (1972): "Metrics and morphophonemics in the Rigveda", In: *Contributions to generative phonology*, M. Brame (ed.), 171-200, Austin: University of Texas Press.
- \_\_\_\_\_. (1974): "On the evaluation measure", In: *Papers from the parassession on natural phonology*, A. Bruck, R. Fox and M. LaGaly (eds.), 328-37, Chicago Linguistic Society.

- Ohala, M. (1974): "The abstractness controversy: experimental input from Hindi", Lg. 50, 225-35.
- Sherzer, J. (1970): "Talking backwards in Cuna: the sociological reality of phonological descriptions", Southwest Journal of Anthropology 26, 343-53.
- Zeps, V. (1963): "The meter of the so-called trochaic Latvian folksongs", International Journal of Slavic Linguistics and Poetics, 7, 123-8.
- \_\_\_\_\_. (1973): "Latvian folk meter and styles", In: A Festschrift for Morris Halle, S. Anderson and P. Kiparsky (eds.), 207-11. New York: Holt.



## THE PSYCHOLOGICAL REALITY OF WORD FORMATION AND LEXICAL STRESS RULES

Anne Cutler, Experimental Psychology, University of Sussex, England

Introduction

'Psychological reality' has both a strong and a weak sense. In the strong sense, the claim that a particular level of linguistic analysis X, or postulated process Y, is psychologically real implies that the ultimately correct psychological model of human language processing will include stages corresponding to X or mental operations corresponding to Y. The weak sense of the term implies only that language users can draw on knowledge of their language which is accurately captured by the linguistic generalisation in question. For certain linguistic constructs, this weak sense embodies no more than a claim to descriptive adequacy; for example, the intuitions which the weak reading of 'psychological reality of the phoneme' predicts speakers will show are the same distributional data which led to the postulation of such a construct in the first place. This is not true of transformational rules - even to claim the weak sense of psychological reality for these is to claim that speakers can draw on knowledge at some level of the structures preceding and following application of the rule.

Lexical stress rules and word formation rules are transformational in nature. Within the grammar, the former are generally assumed to comprise part of the phonology, whereas the latter are claimed by some (Aronoff 1976) to constitute a separate stage preceding application of all phonological rules.

I wish to argue that the available evidence suggests psychological reality in the weak sense for both types of rule, as currently formulated in linguistic theory, but psychological reality in the strong sense for neither. (Note that this argument cannot be generalised to other phonological descriptions; see Fromkin (1973) for an argument in favor of strong psychological reality of abstract phonological representations).

Lexical Stress Rules

I have previously argued (Cutler 1977) that speech error evidence does not suggest the application of lexical stress rules in the production process, i.e. that lexical stress errors do not exemplify the misapplication of stress rules. What might we expect from an error in stress rule application? Fay's (1977a) argument for the strong psychological reality of syntactic transformations

is based on errors which Fay claims show that a particular rule (a) has failed to apply (what he said? for what did he say? is analysed as failure to apply Subject-Auxiliary Inversion), or (b) has applied only partially (Do I have to put on my seat belt on? is explained as application of the movement but not the deletion involved in Particle Movement). Since the function of lexical stress rules is to assign greater relative prominence to one syllable in a word than to others, one might expect that either failure to apply the appropriate rule or only partial application would result in less than the expected difference in degree of prominence between the syllables of a word. That is, if no stress rule applied at all one might expect all vowels in the word to be (equally) prominent, or, possibly, (equally) non-prominent; if, say, the Stress Adjustment Rule failed to apply one might expect a syllable to bear tertiary stress when it should be unstressed, etc.<sup>1</sup> But in fact lexical stress errors result always in primary word stress falling on the wrong syllable, not in lack of differentiation between syllable stress levels. Failure to apply the Alternating Stress Rule (Chomsky and Halle 1968: 78) would indeed result in stress falling on a wrong syllable, e.g. the third syllable of nightingale; but my corpus of lexical stress errors contains not a single such example.

A more complicated hypothesis could be proposed in which, for example, final consonants were misidentified, or the syllables in the word counted wrongly, so that stress ended up on the wrong syllable. But this hypothesis, like the hypothesis that a rule has not applied, in no way predicts the most striking characteristic displayed by lexical stress errors. This is that the syllable which wrongly bears stress is always a syllable which bears stress in another word with the same item. Typical errors are: economist (cf. economic); photographing (cf. photography); conflict<sub>N</sub> (cf. conflict<sub>V</sub>); disadvantageous (cf. disadvantage).

An explanation of these errors which does account for this curious regularity is the following: derivationally related words are in some sense stored together in the mental lexicon, with each word's individual specification including inter alia an indication of stress pattern (stressed syllable); a stress error occurs when

- 
1. Such errors do occur, but only when another word derived from the same base has the intruding stress pattern; e.g. [djúpIkèt] for [djúpIkət].

the stress syllable marking selected is not the one belonging to the target word, but that belonging to one of the other words in the group. (This explanation also accounts for the second, corollary, regularity exhibited by stress errors: they occur only in derived words and only in members of the Latinate section of the English vocabulary. The Germanic section of English is much less rich in morphologically related pairs of words with different syllables stressed, hence it provides less often the necessary conditions for occurrence of a lexical stress error).

It is clear that this explanation, by assuming stress pattern to be marked in the lexicon, implies that lexical stress rules do not apply in the course of language production.

However, there would seem to be no doubt that English speakers can draw on knowledge about the principles governing stress assignment in their language. Many experimental studies (e.g. Ladefoged and Fromkin 1968; Trammell 1978) have found that subjects' pronunciations of non-words or unfamiliar words conform fairly well to the predictions of the lexical stress rules; although Nessly (1977) used similar data collection methods to adduce evidence in favor of his own version of the rules rather than Chomsky and Halle's. Since language users normally find little difficulty in the task of assigning lexical stress in unfamiliar words, names and nonsense words, some representation of the principles underlying English stress assignment must be available to them, i.e. something more abstract than the mere aggregate of all the stress markings stored for all the individual words in their lexicon.

#### Word Formation Rules

Aronoff (1976:22, 46) and Halle (1973:16) specifically exclude any claim to psychological reality of word formation rules in the strong sense. Nevertheless there is evidence from speech errors which could be interpreted as favoring such a claim. Admittedly, one hardly ever finds errors in which a word formation rule seems to have failed to apply, i.e. substitution for the target word of the word or morpheme (depending on one's formulation of the rules) which formed the base of the target - say, familiar for familiarity; for one thing, preservation of target form class is one of the strongest characteristics of word substitution errors of any kind (Fromkin 1973; Fay and Cutler 1977). But errors do occur in which the wrong ending, albeit one appropriate to the form class, is produced: derival for derivation (Fromkin 1977), self-indulgement

for self-indulgence. A possible interpretation of these errors is that the wrong word formation rule has been applied.<sup>2</sup>

It will be obvious, however, that the model suggested in the previous section excludes the application of word formation rules in production as firmly as it excludes the application of lexical stress rules; if word formation rules operate, stress could not be marked in the lexicon as it would be dependent on the operation of the word formation rules. Can this model assign an interpretation to the suffix errors mentioned above? One obvious remark to be made about these errors is their similarity to prefix errors as discussed by Fay (1977b). Prefix errors result in one prefixed word substituting for another (e.g. intention for attention) or a non-word being formed by the addition of an inappropriate prefix (concustomed for accustomed). Similarly suffix errors can result in non-words (e.g. likeliness for likelihood) or in words (necessitous for necessary; these latter errors, word substitutions in which target and error differ only in the suffix, are of course difficult to distinguish from semantic errors and malapropisms). Fay suggested that prefixed words with the same stem might be stored together in the lexicon, and a prefix error result when not the target prefix but a neighbouring prefix was selected by mistake. It is clear that a similar proposal could account for suffix errors producing real words. Thus the lexical entry for a word family would be headed by the stem; the detailed entry for each member of the family would specify affixes, if any, number of syllables (see Engdahl (1978)) and an indication of which syllable should bear lexical stress. To account, however, for both prefix and suffix errors which produce non-words, the model needs to be extended, perhaps to allow the production device to select an appropriate affix from its affix inventory in cases in which the target affix became in some way momentarily unavailable. (It is noteworthy that even when an affix error includes a stress error, stress in the error occurs on a syllable which bears stress in some member of the word family.) To propose factors which might precipitate affix unavailability, i.e. which might render the affix temporarily difficult for the production device to interpret, is, however, to enter the realm of pure

2. These errors show no general tendency for affixes with + or # boundaries to prevail, or for more productive affixes to replace less productive.

speculation. It is to be hoped that more light will soon be shed on this issue; for the time being we must acknowledge that the evidence does not strongly support any particular model.

There is no doubt at all, however, that the facts of word formation have a claim to psychological reality in what we have identified as the weak sense. All the speech error evidence which has been discussed above and which has been interpreted as support for a model of the mental lexicon in which related words are stored together also provides clear support for the psychological reality of morphological structure. A considerable body of psycholinguistic evidence also supports this conclusion (e.g. Taft and Forster 1975). Whether or not rules of word formation of the particular type proposed by Aronoff are available to English speakers to generate new and nonce words is however uncertain. Aronoff and Schvaneveldt (1978) report that subjects in a lexical decision study are more likely to produce false positive responses to non-words formed with the productive suffix -ness than with the less productive suffix -ity, a result predicted by Aronoff's model.

However the results of an informal study of my own were less clearcut. In this study subjects were asked to choose between two candidates for words to fill what amounted to a gap in the language (e.g. to choose between excusal and excusement for 'act of excusing'); each pair of neologisms comprised one word formed with a # boundary (-ness, -ment, -ise, -ish, -y) and another formed with a + boundary suffix (the latter, which often result in stress falling on the suffix rather than on the stem, are considered to be less productive than the # boundary suffixes). Many of the words used were listed in the OED, but none in the Concise Oxford Dictionary, and in fact none of the 12 subjects, graduate students and faculty in psychology and language, claimed to recognise any word.

Since I used only 24 pairs and made no attempt to cover all possible combinations the results can hardly be considered conclusive. Nevertheless some interesting tendencies came to light. In general, subjects showed approximately equal preference for the more and the less productive endings. All subjects preferred excusal to excusement and despisal to despisement, although the OED lists all 4 forms; similarly, subjects preferred amassal and adressal although the OED lists only amassment and addressment. -ness was preferred to -ity for sinister (OED lists both sinisterity and sinisterness for 'quality of being sinister') and incestuous (OED: -ness only),

but accidentality was preferred to accidentalness (OED has both). For verb formations subjects seemed not to be able to make confident choices, and no clear trends emerged; an indication of the confusion can perhaps be seen in the fact that whereas more subjects preferred rapidify to rapidise for 'make rapid', vapidise was chosen more often than vapidify for 'make vapid'. Adjectives revealed yet another pattern of results in that subjects formed two clear groups, those who consistently preferred the less productive + affixes and chose, e.g., spectatorial, plumageous, and dowagerial, and those who consistently chose the more productive # affixes, i.e. spectatorish, plumagy, dowagerish.

The most that can be extracted from these findings is the conclusion that English speakers do not exhibit a great degree of unanimity in their choice of nonce formations. However some light is shed on the psychological reality of word formation processes by a comment made by several subjects independently, namely that although words formed with the + affixes (-al, -ity, -ify, -ial, -ous) were aesthetically more pleasing and would be preferred as permanent additions to the vocabulary, a # affix would generally be more useful to achieve understanding in everyday conversation. Thus although villagerial might in general be preferable to villagerish as an English word, the latter would be more likely to get the message across to an audience not expecting an unfamiliar word. Words with # affixes, which leave stress on the stem, are in other words recognised by speakers to be morphologically more transparent.

#### Conclusion

Morphological structure is psychologically real in that English speakers are aware of the relations between words and can form new words from old. The principles underlying lexical stress assignment are psychologically real in the sense that speakers know the stress pattern of regularly formed new words. The extent to which such knowledge proceeds from competence in the language or awaits conscious insight into morphological relationships is however unclear. It has frequently been suggested to me that morphological influences apparent in my stress error corpus results from error collection within a highly literate and linguistically sophisticated population. If so, then a speaker of English who knows, for example, the words economic and economist but is unaware of any relation between them should presumably not produce a stress error involving either of them. There is certainly no reason why

the structure of the mental lexicon should not be altered as a result of new knowledge about word structure being incorporated in the form of newly set up groupings or connections. But it is also possible that we know more than we are aware of. Recall Fay's discussion of prefixed words; how many of us are consciously aware, for example, that the stem spect in respect appears also in expect? It is at least possible that our mental lexicon could contain such knowledge even if we were not capable of making conscious use of it.

#### References

- Aronoff, M. (1976): Word Formation in Generative Grammar, Cambridge, Massachusetts: MIT Press.
- Aronoff, M. and R. Schvaneveldt (1978): "Testing morphological productivity". Unpub. MS.
- Chomsky, N. and M. Halle (1968): The Sound Pattern of English, New York: Harper and Row.
- Cutler, A. (1977): "Errors of stress and intonation", paper presented to 12th International Congress of Linguists, Vienna.
- Engdahl, E. (1978) "Word stress as an organizing principle for the lexicon". Papers from the Parasession on the Lexicon, Chicago Linguistic Society.
- Fay, D. (1977a): "Transformational errors", paper presented to 12th International Congress of Linguists, Vienna.
- Fay, D. (1977b): "Prefix errors", paper presented to 4th International Linguistics Meeting, Salzburg.
- Fay, D. and A. Cutler (1977): "Malapropisms and the structure of the mental lexicon", Linguistic Inquiry 8, 505-520.
- Fromkin, V. (1973): Speech Errors as Linguistic Evidence, The Hague: Mouton.
- Fromkin, V. (1977): "Putting the emphasis on the wrong syllable", in Studies in Stress and Accent, L.M. Hyman (ed.), Los Angeles: USC Press.
- Halle, M. (1973): "Prolegomena to a theory of word formation", Linguistic Inquiry 4, 3-16.
- Ladefoged, P. and V. Fromkin (1968): "Experiments on competence and performance", IEEE Transactions on Audio and Electroacoustics, 16, 130-136.
- Nessley, L. (1977): "On the value of phonological experiments in the study of English stress", in Studies in Stress and Accent, L.M. Hyman (ed.), Los Angeles: USC Press.
- Taft, M. and K. Forster (1975): "Lexical storage and retrieval of prefixed words", JVLVB 14 638-647.
- Trammell, R.L. (1978): "The psychological reality of underlying forms and rules for stress", J. Psycholing. Res. 7, 79-94.

PROBLEMS IN ESTABLISHING THE PSYCHOLOGICAL REALITY OF  
LINGUISTIC CONSTRUCTS

Bruce L. Derwing, Department of Linguistics, University of Alberta,  
Edmonton, Alberta, Canada T6G 2H1

The "psychological reality" issue in linguistics - and in phonological theory in particular - has many more facets than seem to be generally recognized. The first problem is the recognition that a problem in fact exists. Under the influence of ideas which developed originally in comparative philology, the prevailing linguistic philosophy has long been one of autonomy: a language has been viewed as a kind of isolated "natural object" which could be investigated independently of the psychology of its speakers and hearers. In recent years, this misapprehension has led to a concept of linguistic "competence" (Chomsky 1965) which consists of nothing more than an arbitrary set of "coding principles" (Straight 1976) abstracted by linguists from linguistic data and treated as something quite distinct from the mechanisms of listening and speaking. Yet, in fact, a language is not an isolated "thing" at all, but is rather the product of various psychological and physiological processes which take place within human beings. Physically, the language product can be studied in the form of speech articulations, acoustic waves, or peripheral auditory events, but in none of these three observable, physical states can we find anything which smacks of linguistic "structure" (not even "phones," which already involve considerable processing by the human perceptual apparatus). Linguistic "structure," therefore, if this term refers to anything real at all, must refer to representations or interpretations imposed upon the speech signal by language users, normally as part and parcel of the communication event itself (Derwing 1973, 302-307). In short, psychological reality is not merely a convenient luxury which linguistic theory may or may not choose to be concerned with, but is rather the sine qua non for any linguistic construct which aspires to anything more than an epiphenomenal or artifactual status, and hence for any linguistic theory which can justifiably claim to go beyond the bounds of an arbitrary taxonomic system.

It is for this very reason, in fact, that all of modern "autonomous" linguistics suffers from an insoluble non-uniqueness

problem: any set of language forms can be correctly (i.e., accurately) described in many different ways, as even the simple example of the English plural inflection clearly shows (Derwing 1974). This implies that pure "linguistic" or "internal" evidence (i.e., evidence about "static" language forms, etc.) is quite inadequate to distinguish a wide range of theoretical alternatives. The only apparent solution to this problem (apart from the adoption of arbitrary principles for grammar "evaluation") is to redefine the nature of the discipline: to say that the goal of linguistics is not merely to describe utterance forms, but rather to describe the knowledge and abilities which speakers have to produce and comprehend them. Linguistic claims now become subject to the test of truth: whereas the forms will admit of numerous possible descriptions, there are many psychological claims about what the speaker knows or does which can be shown to be wrong or inadequate. So an expanded domain of "psycholinguistic" evidence can help to sort out alternatives which the traditional kinds of "linguistic" data could not.

To recognize the need to psychologize linguistics is one thing, however, and the actual practice is something else again. Chomsky himself declared linguistics to be a branch of cognitive psychology a full decade ago (1968), yet he and his followers still continue to embrace many of the same old "pernicious ideas" (McCawley 1976) which prevent this conception from becoming anything more than a slogan. In other words, while the so-called "Chomskyan revolution" may well have entailed a terminological re-orientation in the direction of the psychologization of linguistic jargon and associated claims, no corresponding methodological revolution accompanied these changes, with the result that the generative grammarians "continued to practice linguistics as it has always been standardly practiced" (Sanders 1977, 165). Such linguists may thus claim to seek or establish "psychological reality," yet they still persist in evaluating their theories on the basis of various "simplicity" considerations rather than on the basis of independent psychological evidence (as if the more general theory were, in fact, the most psychologically "valid"; contrast Fromkin 1976, 56, with Steinberg 1976, 385-386.)

But we are still merely scratching at the surface of the problem. It has become commonplace nowadays to find exhortations

in the linguistic literature to "expand the data base," often in much the same directions as outlined above, yet Greenbaum seems quite justified in expressing the doubt "whether linguists will abandon a particular linguistic formulation on the basis of psycholinguistic evidence" (1977, 127). Why should they? For, after all, since most linguistic theorizing was done within the non-psychological or "autonomous" linguistic tradition, it is seldom clear what particular psychological claims, if any, are to be associated with any particular linguistic analysis. Obviously, before we can ever hope to make use of new kinds of evidence to test or evaluate psychological claims, we must first know what the particular claims are that we are required to evaluate.

This is the crux of what I have called the interpretation problem for grammars (Derwing 1974). If grammars merely describe utterance forms, then evidence about such forms is the only kind relevant to the evaluation of grammars, and a selection from among competing grammars can only be made on the basis of criteria which are ultimately arbitrary. But if grammars relate in any way to psychological events or states, then we need to interpret grammars psychologically so as to make it clear what the new empirical implications of these grammars are. In other words, a formal grammar requires a psychological interpretation before it can become part of a psychological theory, and it is only the combination of the grammar plus the interpretation which can be put to an experimental test.

Now the problem of interpretation is not nearly so severe with respect to some of the older, more concrete linguistic notions as in the case of many of the more recent, abstract developments. In Derwing & Baker (1977), for example, a summary is provided of various straightforward interpretations, relevant tests, and new experimental data which help to answer the question of which, if any, of several obvious ways of describing the English plural inflection is psychologically real. Serious problems arise, however, when we come to analyses of the type discussed in Anderson (1974, 54-61), which involve the positing of single "underlying" lexical representations and "extrinsically ordered" phonological rules. Even ignoring the major problem of what psychological interpretation to place upon the general notion of the grammatical "generation" of forms (cf. Crystal 1974, 303), what psychological

sense can possibly be made, first of all, of a notion of rule "ordering" which has no relation to real time? To my knowledge, no one has ever even proposed a sensible real-world analogue of this idea, and without an interpretation, to repeat, it is impossible to tell what kind of experimental test is even relevant to evaluate its validity. Fortunately, in this instance, at least, the concept is one that linguistic theory seems to be able to get along very nicely without, merely by reformulating all rules in such a way that no arbitrary ordering relations are required among them (cf. Derwing 1973 & 1975). But we are still left with the problem of what psychological content we can associate with the linguist's notion of the "underlying" or "base" form in phonology. A few suggestions have at least been made in this case (e.g., Linell 1974; Ingram 1976; Birnbaum 1975), but none of them have yet seemed compelling enough for anyone to risk taking one out onto an experimental limb. There is, in any event, another, less direct route which can be taken in connection with this particular evaluation problem. The keystone argument is that there is no basis for positing a single "underlying" lexical representation for any set of supposed "morpheme alternants" unless the alternants in question can indeed be shown to represent the "same morpheme" for speakers. Thus a test which assesses a speaker's ability to "recognize morphemes" can indirectly provide evidence relevant to the question of the extent to which psychological theories might plausibly be constructed which incorporate the linguistic notion of the "underlying" form. For example, on the basis of "morpheme recognition" data collected by means of tests described in Derwing (1976), there is reason to believe that typical speakers judge a word-pair to contain a common morpheme only if the two words involved share a certain "critical" degree of both semantic and phonetic similarity, as independently assessed (Derwing & Baker in press). On this evidence, therefore, any linguistic analysis which posits a common lexical representation for words such as fable and fabulous, which lie outside of this "critical" area, is not even psychologically feasible for more than a very small minority of speakers.

While the recognition and solution of the interpretation problem represent, I think, the main barrier to the establishment of the psychological reality of linguistic constructs, there are still



quite a number of smaller obstacles which also have to be faced and overcome. For one thing, we must learn to resist the temptation to be "bathtub experimentalists" (i.e., prone to the cry of "Eureka!"). For even an investigator who fully recognizes the need both to interpret and to test linguistic theories on psychological territory may well (for lack of laboratory experience, for example) fail to anticipate many of the difficulties which can arise out of the very activity of devising, carrying out, and finally evaluating experiments. The most insidious of these difficulties, no doubt, is the one associated with the experimental artifact. For just as (autonomous) linguistic theorizing has yielded many concepts which have no real-life analogues in the knowledge or skills of real language users, so a particular experimental technique can also yield data which are more representative of the technique (or of his subjects' ingenuity) than of the subjects' control of the phenomenon of interest. A particular experiment does not always test in practice what the experimenter thinks it is testing in theory. I have encountered this problem at least twice in my own research (cf. Derwing 1976, 43-50) and Fromkin (1976) properly takes a few experimenters to task for perhaps jumping too fast to conclusions because of it. But in the last analysis there is only one sure way to dispel doubts about the "experimental artifact" and that is via the very painstaking route of cross-methodological verification: each evaluation problem must be approached by means of a variety of alternative experimental routes, in order to insure that the results obtained are independent of any particular experimental procedure.

There are, of course, other methodological problems to be mentioned, as well. There is always, for example, the possibility of the "just plain goof" whenever experimental data are collected, interpreted and evaluated, a danger that springs from causes as trivial as the mispunching of data cards to others as abstruse as failure to attend to assumptions which underlie a particular statistical model. Yet the most common type of error to sneak through a data analysis unattended, perhaps, is the one that results from a failure to take due cognizance of uncontrolled confounding variables (cf. Derwing & Baker 1977, 100), with the result that one's interpretation may be based on an apparent cause rather than the real one. But, again, there is no sure or simple formula to

guarantee safe passage through such treacherous and unpredictable waters as these; one can only take the utmost care possible in his own work, then hope that his readers and critics will pick out whatever errors and oversights may remain.

Finally, there is also the problem of the extraneous or "nuisance" variable, so called, no doubt, because it is often so very hard to eliminate from the experimental situation, even when the investigator may know full well that it is there. In my own "morpheme recognition" research, for example, the interpretation of the data is continually muddled by the factor of possible orthographic interference. How much are "linguistic intuitions" conditioned by the academic task of learning how to read, thereby complicating our efforts to understand the "natural" course of language acquisition through mere exposure to spoken language forms under normal circumstances of use? (A very similar question is the one concerning the very validity of the "linguistic intuitions" of subjects who have already been exposed to any significant degree of formal linguistic training; cf. Derwing in press.) Answers to such questions can only be partially and very tentatively answered so long as one is forced to deal with literate (or "non-naive") experimental subjects. I am very happy to see, therefore, that some aspects of my work are soon to be replicated and extended to the study of Lapp morphology by R. Endresen of the University of Oslo, for included in his population samples will be many speakers who are not only linguistically untrained, but also illiterate in their own language, thereby making it possible to investigate systematically at least some effects of the orthographic variable. Unfortunately, not all "nuisance" factors can be so conveniently dealt with, and these others will continue to constitute one of the more troubling aspects of trying to advance our knowledge by means of controlled experimental research. But since this is the way of science and the only secure route we know of for establishing knowledge about the world and its inhabitants, we have little real choice but to face them all head on.

#### References

- Anderson, S.R. (1974): The organization of phonology, New York: Academic Press.
- Birnbaum, H. (1975): "Linguistic structure, symbolization, and phonological processes", in Phonologica 1972, W.U. Dressler et al. (eds.), 131-143, Munchen & Salzburg: Wilhelm Fink.

- Chomsky, N. (1965): Aspects of the theory of syntax, Cambridge, Massachusetts: MIT Press.
- Chomsky, N. (1968): Language and mind, New York: Harcourt Brace & World.
- Crystal, D. (1974): Review of R. Brown, A first language. Journal of Child Language 1, 289-334.
- Derwing, B.L. (1973): Transformational grammar as a theory of language acquisition, London: Cambridge University Press.
- Derwing, B.L. (1974): "English pluralization: a testing ground for evaluation", to appear in Experimental linguistics, G.D. Prideaux et al. (eds.), Ghent: E. Story-Scientia.
- Derwing, B.L. (1975): "Linguistic rules and language acquisition", Cahiers Linguistiques d'Ottawa, No. 4, 13-41.
- Derwing, B.L. (1976): "Morpheme recognition and the learning of rules for derivational morphology", Canadian Journal of Linguistics 21, 38-66.
- Derwing, B.L. (in press): "Against autonomous linguistics", in Evidence and argumentation in linguistics, T. Perry (ed.), Berlin & New York: de Gruyter.
- Derwing, B.L. & W.J. Baker (1977): "The psychological basis for morphological rules", in Language learning and thought, J. Macnamara (ed.), 85-110, New York: Academic Press.
- Derwing, B.L. & W.J. Baker (in press): "Recent research on the acquisition of English morphology", in Studies in language acquisition, P.J. Fletcher et al. (eds.), London: Cambridge University Press.
- Fromkin, V.A. (1976): "When does a test test a hypothesis?", in Testing linguistic hypotheses, D. Cohen et al. (eds.), 43-64, New York: Wiley.
- Greenbaum, S. (1977): "The linguist as experimenter", in Current themes in linguistics, F.R. Eckman (ed.), 125-144, New York: Wiley.
- Ingram, D. (1976): "Phonological analysis of a child", Glossa 10, 3-27.
- Linell, P. (1974): Problems of psychological reality in generative phonology, Uppsala University, Department of Linguistics.
- McCawley, J.D. (1976): "Some ideas not to live by", Die neueren Sprachen 75, 151-165.
- Sanders, G.A. (1977): "Some preliminary remarks on simplicity and evaluation procedures in linguistics", in L.G. Hutchinson (ed.), Minnesota Working Papers in Linguistics and Philosophy of Language, No. 4, 155-167.
- Steinberg, D.D. (1976): "Competence, performance and the psychological invalidity of Chomsky's grammar", Synthese 32, 373-386.
- Straight, H.S. (1976): "Comprehension versus production in linguistic theory", Foundations of Language 14, 525-540.



ARGUMENTS AND NON-ARGUMENTS FOR NATURALNESS IN PHONOLOGY:  
ON THE USE OF EXTERNAL EVIDENCE

Wolfgang U. Dressler, Institut für Sprachwissenschaft,  
University of Vienna, Austria

§1.0 The concept of naturalness has become a major concern for many phonologists. In my view, the concept of naturalness should be best regarded as a basic principle of a phonological theory and should be tested by the judicious use of external (or substantive) evidence.

As to the relationship of naturalness to psychological reality, my point is that a natural phonological analysis of a phenomenon claims psychological reality for its concepts and constructs. However, not all psychologically real constructs in a phonological analysis need to be phonologically natural. E.g., a phonological process (henceforth PR) posited by the linguist may refer to constructs of natural morphology (cp. Mayerthaler to appear), especially in case of so-called morphological rules (cp. Dressler, 1977a).

§1.1 In the theory of Natural Phonology (henceforth NatPhon), as proposed by Stampe since 1968 (see now Donegan and Stampe to appear) and 'Polycentristic Phonology' (Dressler 1977a), naturalness occupies a central place. Phonological systems are phonetically (and I add, psychologically and, to a lesser degree: sociologically, historically) motivated. The basic constructs of Natural Processes in the sense of "mental substitutions which systematically but subconsciously adapt our phonological intentions to our phonetic capacities" (Donegan and Stampe to appear, §1, including its perceptive converse) are substantive universals.

§1.2 Similar to adherents of NatPhon, S. Schane and M. Chen (see Sommerstein 1977, 230, 233) have claimed that particular languages select PRs from a fixed universal set of natural processes and may impose constraints on their applicability. In the best of cases a PR forms a subset of a universal process (as characterized by the theory) and any restrictions vis-a-vis the general form of the respective universal process can be derived from the hierarchies of the universal process and from a fairly small number of principles of restrictions.

But what if a PR is not such a regular subset of a universal process? In this case NatPhon (or at least Polycentristic Phonology) cannot appeal to frequency or intuitive plausibility, but

must explain why the given PR is not a regular subset of a universal process. Several avenues are open: 1) Modification of the universal process. 2) The deviation is due to language acquisition; in this case well-motivated linguistic and psychological concepts must explain the deviancy. 3) The deviation is due to historical circumstances (including sociological factors). 4) The PR is not totally (phonologically) natural, a possibility avoided in NatPhon (but cp. Dressler 1977a; Sommerstein 1977, 235f). Since such PRs (diachronically) must go back to totally natural PRs, explanation 4) includes explanations 3) and 2).

§1.3 Thus, it becomes clear that external evidence, at least from language acquisition, diachrony, and sociolinguistics is not external for NatPhon, but forms an integral part of the area it has to cover. Moreover, there is no theoretical or methodological principle which should exclude other dimensions of external evidence from investigation:

§1.3.1 Take sociophonology: The restriction to the investigation of only one level of formal, maximally differentiated speech as practised in most of generative phonology and almost all of structural phonology is an undue limitation of interest and of access to natural speech, whose variation is apt to give important insights even to formal principles of rule application (cp. Dressler 1975). However, any detailed and theoretically sound work on casual vs. formal speech presupposes the inclusion of both, psychological/psycholinguistic theory (cp. Vanecek and Dressler 1977) and sociological/sociolinguistic theory (cp. Wodak and Dressler 1978).

§1.3.2 Or: The differential (and always non-random!) breakdown of phonology in aphasia gives important insights into the structure of phonology. However, studies so far have not completed the desirable integration of all disciplines relevant to aphasia, e.g. the brilliant thesis of Keller (1975) neglects all recent phonological theories, whereas the present writer's studies (Dressler 1977b; 1978) have not yet integrated neuropsychology. For other types of external evidence, see Linell (1974), Fischer-Jørgensen (1975, 224ff), Zwicky (1975), Skousen (1975).

## §2. Non-arguments for naturalness

In the literature we find certain non-arguments/fallacies:

§2.1 "Facts about the real working of the brain are most important".

Anttila (1977, 221) believes to have found direct evidence, as opposed to indirect neurolinguistic evidence, against generative grammar, when he cites the biologist W. Wieser about the brain not working exactly, often blundering and correcting itself, not proceeding logically, but according to similarities, being extremely redundant, etc. However, Wieser has informed me that these phenomena at the micro-level do not preclude precise rules at the macro-level (which is the level of interest for linguistics), just as Heisenberg's indeterminacy relation does not vitiate the precise working of laws of classical physics in macrophysics. Here we might speak of a micro-anatomic fallacy.

§2.2 There is a similar fallacy which one might call the macro-anatomic fallacy or mistaken equation of phonology and phonetics, which is an exaggeration of the Physical/Phonetic Basis Condition (Botha 1978 II, 16ff) of phonology. This line of argument neglects the interaction between phonological and morphological or phonological and lexical naturalness (cp. Dressler 1977a) and of what Hyman (1977) has called phonologization (which in my view starts with allophonic PRs producing extrinsic instead of intrinsic allophones).

§2.3 Still more common is the false equation of naturalness with concreteness, since as a result of refusing the abstractness involved in standard generative phonology, many phonologists have regarded concreteness as a virtue in itself. However, phonological concreteness has often been achieved at the expense of morphology for which very few 'concrete phonologists' (e.g. Skousen 1974) have cared to provide a theoretical framework. More important still, concreteness has been defined (if at all) as restrictions on the relationship between underlying phonological and surface phonetic representations. In my opinion it is possible to define the naturalness of processes and of representations (be it as structural symmetries as found by phonemicists or natural asymmetries as derived from processes, see Stampe (1973)), but not the naturalness of relationships between representations. Notice both the failure of Kenstowicz and Kisseberth (1977) to find universal formal constraints on the distance between phonological and phonetic representations (cp. Gussman 1978, 154, 167f; Sommerstein 1977, 237 n. 47), and the undesirable results of the much more rigorous restrictions of Natural Generative Phonology (see Hooper (1976) and

its critique by Gussman (1978, chapter 1) and Donegan and Stampe to appear, §3.1., §4).

As an example I simply want to refer to the abstract analysis of German [ŋ] as underlying /ng/ (discussed in detail in Dressler to appear; cp. Dressler (1977a, 51)). For the much debated PR  $g + \emptyset/\eta$  — (except before non-centralized vowel), I have found external evidence, e.g. in loan-word integration and sociophonological variation (e.g. [<sup>l</sup>ʌŋgɛla] vs. [<sup>l</sup>ʌŋɛla] 'Angela'), in child language (Mandarine 'tangerine' + [mʌŋgʌ'ri:nə] vs. [mʌŋɛ'ri:nə]) and aphasia (see Stark 1974). Thus, multiple external evidence has been found in support of the psychological reality of this PR, although this analysis implies a very abstract underlying representation (cp. Kenstowicz and Kisseberth (1977, 7f, 53), Gussman (1978, 168), see below §3.4).

§2.4 Often natural is falsely equated with productive. This equation (Fischer-Jørgensen (1975, 228f); Skousen (1975); Linell (1976), etc.) might hold most of the time, but not always (Dressler (1977a, 1977c)).

§2.5 Still weaker and never explicitly justified is the equation of natural and (e.g. typologically) frequent. Frequency might be a first indicator for the phonologist looking for universals, but what counts is explanation in the sense of causal argumentation.

### §3. Counterarguments against external evidence

§3.1 "External evidence is unnecessary, internal evidence suffices". This 'Nonnecessity Thesis' has been proved by Botha (1978 II) to be incompatible with empirical mentalism. Formal, 'pure' linguistics cannot alone do the job of vouching for psychological reality. Due to the serious underdetermination of standard data (internal evidence), various sources of external evidence must be adduced (cp. §1.3 and Botha (1978 II, III §5.3)).

§3.2 "External evidence is too unclear". However, internal evidence based on intuitions as utilized in generative phonology is unclear itself in many respects as Ringen (1975) has shown. Moreover, it must be noted that evidence from diachrony and loan-word integration seems to be accepted by many who shun other external evidence.

Unfortunately, the use of both types of evidence has been grossly simplified by most generative phonologists; for loan-words see Fischer-Jørgensen (1975, 229), Kiparsky (1973, 112ff), Dressler

(1977a, 35ff). As to diachrony, both structuralists and generativists have limited themselves far too often to nomological explanations (e.g. symmetry, rule simplification), while neglecting the all-important genetic explanation, e.g. by confusing sound change with sound correspondences; thus context-free processes have been liberally adduced as evidence, although they are, I believe, always the final result of generalizing context-sensitive sound change.

§3.3 External evidence shows "what in fact counts as internal evidence" (Kenstowicz and Kisseberth 1977, 3). Does this mean that e.g. English loan-words in Japanese might be used to demonstrate the necessity of morpheme structure constraints or redundancy rules within phonological theory, but not for corroborating their specific forms in Japanese itself?

§3.4 "Internal evidence is more important than external evidence", a view held by many (called the Nonprivileged Status Thesis by Botha (1978 II, 12f)). However, quite apart from its theoretical shakiness (cp. §1), there are counterexamples: E.g. the abstract analysis of English [ŋ] as /ng/ rests on exceptional (and thus suspect) alternations like lo[ŋ], lo[ŋ]est, whereas the normal, productive superlatives are e.g. bori[ŋ]est, winni[ŋ]est; but external evidence for the abstract analysis is excellent (starting with Fromkin (1973, 223)). Even more extreme is the German situation, where in most varieties internal evidence is restricted to distributional evidence (Vennemann 1970), which generativists usually esteem much less than evidence from alternations, alternations in this case exist only in external evidence (see above §2.3).

§3.5 Botha (1970, 130ff) has deplored the 'qualitative type jump' from internal to external evidence and the lack of criteria of adequacy. Since then he has revised his standpoint and has demanded the construction of "bridge theories" mediating between linguistics and other disciplines relevant for the given type of external evidence (Botha 1978 III, 27ff). But 'hyphenated' disciplines, such as psycholinguistics, sociolinguistics, neurolinguistics have strived just for that since many years!

§3.6 "External evidence is often divergent and incoherent" (Gussman (1978, 167f) happily cites Dressler (1977d, 224), where higher standards in the use of external evidence are demanded). Here Botha (1978 III, 30, 27ff) correctly states "that the relative weight of a given kind of external evidence is a function of the

adequacy of a particular bridge theory". In other cases conflicting external evidence may force us to revise phonological theory (e.g. in the case of introducing Korhonen's concept of 'quasi-phonemes' in Dressler (1977a, 52ff).

§3.7 In connection with §3.6 I want to discuss a problem which seems to strike a heavy blow to the theory espoused here: I have linked naturalness firmly with the universality of natural processes. However, processes actually studied, show different hierarchies, both typologically and in external evidence (cp. Drachman 1977; Ferguson 1978), although hierarchies have been claimed to be an integral part of the universal processes constructed by NatPhon. Whereas Atomic Phonology has found a purely formal solution (criticized by Donegan and Stampe (1977)) to this problem, I want to come back to §1.2. The reactions of an individual to innate physiological and psychological restrictions are determined both by maturation and social environment. In this way I agree neither with (rather mystical) strong claims about innate universals (as in certain quarters of TG), nor with the arbitrariness of the outcome of societal constraints (as implied in marxist critiques of TG). Therefore (in Dressler 1977a) I have spoken only of universal tendencies (one type being universal processes) which necessarily conflict and must be compromised by the language learner: Thus, certain universal processes are suppressed either in the language as a whole or in certain domains of external evidence; or they are restricted in ways allowing different process hierarchies. Moreover, a typology of phonological processes must consider advances made in the theory of typology: e.g. ordering typologies may be multilinear (with branchings).

#### References

- Anttila, R. (1977): Rev. of Linell 1974, Lingua 42, 219-222.
- Botha, R. (1970): The justification of linguistic hypotheses, The Hague: Mouton.
- Botha, R. (1978): On the method of mentalism, Ms., University of Stellenbosch.
- Donegan, P.J. and D. Stampe (1977): "On the description of phonological hierarchies", in CLS Book of Squibs, S. Fox et al. (eds.), 35-38.
- Donegan, P.J. and D. Stampe (to appear): "The study of natural phonology", in Current Approaches to Phonological Theory D. Dinnsen (ed.), Bloomington: Indiana University Press.
- Drachman, G. (1977): "On the notion 'Phonological Hierarchy'", in Phonologica 1976, W. Dressler and O. Pfeiffer (eds.), Innsbrucker Studien zur Sprachwissenschaft, 85-102.
- Dressler, W. (1975): "Methodisches zu Allegroregeln", in Phonologica 1972, W. Dressler and F. Mareš (eds.), Munich: Fink, 219-234.
- Dressler, W. (1976): "Tendenzen in kontaminatorischen Fehlleistungen", Sprache 22, 1-10.
- Dressler, W. (1977a): Grundfragen der Morphonologie, Vienna: Österreichische Akademie der Wissenschaften.
- Dressler, W. (1977b): "Morphological disturbances in aphasia", Wiener linguistische Gazette, 14, 3-11.
- Dressler, W. (1977c): "Morphologization of phonological processes", in Linguistic Studies presented to J. Greenberg, A. Juillard (ed.), Saratoga: Anma Libri, 313-337.
- Dressler, W. (1977d): Rev. Skousen 1975, Lingua 42, 223-225.
- Dressler, W. (1978): "Phonologische Störungen bei der Aphasie", Badania lingwistyczne nad afazją, Warsaw: Ossolineum, 11-22.
- Dressler, W. (to appear): "External evidence for an abstract analysis of the German velar nasal", in Phonology in the 1970's D. Goyvaerts (ed.), Ghent: Story-Scientia.
- Ferguson, Ch.A. (1978): "Phonological processes", in Universals of Human Language, J. Greenberg (ed.), Stanford Univ. Press, 403-442.
- Fischer-Jørgensen, E. (1975): "Perspectives in Phonology", Annual Report of the Institute of Phonetics, University of Copenhagen 9, 215-235.
- Fromkin, V. (1973): (ed.) Speech errors as linguistic evidence, The Hague: Mouton.
- Gussman, E. (1978): Explorations in abstract phonology, Univ. of Lublin.
- Hooper, J. (1976): An introduction to natural generative phonology, New York: Academic Press.
- Hyman, L. (1977): "Phonologization", in Linguistic Studies presented to J. Greenberg II, A. Juillard (ed.), Saratoga, 407-418.
- Keller, E. (1975): Vowel errors in aphasia, Univ. of Toronto.
- Kenstowicz, M. and Ch. Kisseberth (1977): Topics in phonological theory, New York: Academic Press.
- Kiparsky, P. (1973): "Phonological representations", in Three dimensions of linguistic theory, O. Fujimura (ed.), Tokyo: TEC, 1-136.
- Linell, P. (1974): "Problems of psychological reality in generative phonology", Reports from Uppsala University, Department of Linguistics 4.
- Linell, P. (1976): "Morphonology as part of morphology", Phonologica 1976, 9-20.
- Mayerthaler, W. (to appear): Morphologische Natürlichkeit, Ms. TU Berlin.

- Ringen, J. (1975): "Linguistic facts", in Testing linguistic hypotheses, D. Cohen and J. Wirth (eds.), Washington: Hemisphere Public Corp., 1-42.
- Skousen, R. (1974): "An explanatory theory of morphology", Papers from the Parasession on Natural Phonology, CLS, 318-327.
- Skousen, R. (1975): Substantive evidence in phonology, The Hague: Mouton.
- Sommerstein, A. (1977): Modern phonology, London: Arnold.
- Stampe, D. (1973): "On chapter nine", in Issues in phonological theory, Kenstowicz and Kisseberth (eds.), The Hague: Mouton, 44-52.
- Stark, J. (1974): "Aphasiological evidence for the abstract analysis of the German velar nasal", Wiener linguistische Gazette 7, 21-37.
- Vanecek, E. and W. Dressler (1977): "Untersuchungen zur Sprechsorgfalt als Aufmerksamkeitsindikator", Studia Psychologica 19, 105-118. (A preliminary version in Wiener linguistische Gazette 9, 1975).
- Vennemann, Th. (1970): "The German velar nasal", Phonetica 22, 65-82.
- Wodak, R. and W. Dressler (1978): "Phonological variation in colloquial Viennese", Michigan Germanic Studies 4, 1, 30-66.
- Zwicky, A.M. (1975): "The strategy of generative phonology", Phonologica 1972, Dressler and Mareš (eds.), 151-168.

## ABSTRACT PHONOLOGY AND PSYCHOLOGICAL REALITY

Edmund Gussmann, Institute of English, Maria Curie-Skłodowska University, Lublin, Poland

For a number of years now abstract phonological descriptions have come under attack from two different but often related quarters.<sup>1</sup> Firstly, it has been claimed that even within the broad framework of standard generative phonology less abstract solutions are often available; reinterpretations of the data have been achieved by suggesting that certain putative phonological contrasts are in fact morpho-lexical generalisations, i.e. morphologically and lexically rather than phonologically conditioned. Re-analysis or change of underlying representations has also been offered as a viable alternative to manipulating abstract segments and opaque rules. Finally, various modifications in the rule component have been shown to lead to less drastic departures from phonetic representations than those called for by (relatively) abstract positions. The drive towards concreteness seems to have culminated in the rise of so-called 'natural generative phonology' of Vennemann, Hooper and others although a whole range of more or less abstract views has continued to exist; in fact these radically concrete positions are coming under attack now even from those linguists who generally favour concreteness in phonology (cf. Goyvaerts 1978, 125-133). In any case, the type of criticism of abstract solutions that is normally based on evidence internal to the structure of the language cannot be meaningfully discussed without taking into account the grammar as a whole, and this is obviously precluded here. It can be safely assumed that less abstract solutions will be acceptable even to those linguists who favour abstractness in phonology if it can be shown that abstract interpretations are not necessary, i.e. that either the required generalisations can be made without recourse to the abstract machinery or else that the generalisations are in fact wrong and must be replaced by others. It is perhaps worth stressing that in order to evaluate such arguments and counter-arguments one must consider not just individual pairs of rules but rather the phonology as a whole; there has been far too much specula-

-----

(1) The bibliography of the subject is vast and would require several pages. In this report I have restricted myself to just a few items which are directly relevant to the discussion.

tion based on scattered examples and even on inaccurate data.

The other line of attack on abstract positions has involved external evidence which has come to be known as substantive evidence. It has been claimed that the generalisations captured in abstract descriptions are not those that speakers of the language make, i.e. that the abstract generalisations are, in a nutshell, a figment of the linguist's imagination devoid of any psychological reality. This line stresses the need to go beyond the structural facts of the language in search of support for true generalisations. Substantive evidence for such psychologically real regularities has been sought in historical change, the treatment of borrowings, in language acquisition and language loss (aphasia), metrics, dialectal variation, speech errors, secret languages as well as in direct phonological experiments (see Fischer-Jørgensen 1975, 290ff and Zwicky 1975 for good surveys). These are important findings which certainly cannot be overlooked by anybody seriously concerned with psychologically real phonology. They must, however, be handled with extreme caution given the present understanding of the ways in which language is actually used since, as was judiciously observed by Dressler (1977, 224), "the more modalities of external evidence one uses, the more divergent and incoherent results one gets". Let me consider just a few cases.

Polish has a general and typologically very natural rule of devoicing obstruents word finally. In actual speech one often finds that the rule is suspended in certain cases, e.g. in regularly used foreign words and names whether completely assimilated into the language or not - gro[g] 'grog' rather than gro[k], ko[d] pocztowy 'postal code' (in spite of the fact that [d] precedes a voiceless plosive!), possibly because the unvoicing would produce here the humorous kot pocztowy 'postal cat'; in native words it is also suspended for a variety of reasons as in dó[b] '24 hrs., gen.pl.', where the unvoicing would produce a somewhat improper word. Surely no one would like to conclude from such examples that terminal unvoicing is not a psychologically real rule in Polish. Generally speaking, foreign words exhibit specific properties, and most schools of phonology have reflected this fact in one way or another (in addition it seems that one should also recognise varying degrees of foreignness). The fact that some foreign or occasional native words (including, possibly, nonsense words) do not appear

to have undergone a rule cannot be taken as direct evidence for the non-reality of the rule.

Historical evidence, one of the most important sources of substantive evidence, is notoriously difficult to handle in that the paucity or lack of reliable and unambiguous data is not the only factor hampering definite conclusions; any interpretation of change for purposes of verifying general theoretical claims involves assumptions about the mechanisms of change which themselves are not well understood and it also involves assumptions about e.g. the interface between the rules of morphology and those of phonology which is likewise largely unexplored. In view of these problems it is not surprising that examples can be found in the literature purporting to justify both abstract and concrete positions by use of such evidence. The metric evidence available from the works of Kiparsky, Anderson and others seems to support the level of remote representations although, given the variety of theoretical machinery accessible to current linguistic thinking, alternatives could presumably be found.

Slips of the tongue have figured prominently as the window to psychologically real grammars, and Fromkin's (1971) seminal paper has stimulated a lot of interest in this area. Some of her evidence has now become part of the stock-in-trade of those arguing for abstract regularities as, for example, the celebrated case for /ng/ as underlying the phonetic [ŋ]. It would be easy for somebody trying to defend abstract phonology to claim that if /ng/ underlies [ŋ] in a psychologically real sense, then speakers of English must have at their disposal means of arriving at the abstract solution given the data internal to the language. These means could then be generalised to cases where no external evidence can be adduced; this is the position adopted by Kenstowicz and Kisseberth (1977) who incidentally find that the case of the English velar nasal violates all of their constraints on the abstractness of underlying representations. Such evidence is intriguing, but supporters of concrete phonology could easily dispose of it by viewing the slips as resulting from the influence of spelling or something else. I would like to further emphasise, however, that important as such evidence may be, it is not obvious whether much use can be made of it until more is known about the interaction of linguistic knowledge and language use. In our particular case we need some sort

of theory of speech errors against which we could evaluate individual instances for their linguistic significance since one frequently observes not only slips of the tongue that can be shown to reveal something about the underlying reality of language but also instances of errors that appear to make "no sense" linguistically. It is also worth mentioning that different areas often provide contradictory evidence (cf. also Dressler's remark quoted above). The following might be a possible example: slips of the tongue adduced by Fromkin appear to suggest that affricates should be treated as single segments phonetically in English. On the other hand, optional low phonetic rules frequently simplify affricates to spirants in certain contexts so that French and orange end in [ʃ] and [ʒ]. This, of course, could be interpreted as a change in the feature /cont/ but since one also finds the deletion of alveolar plosives in such words as rents, sounds, it seems more plausible to treat both these changes as cases of deletion of the plosive between a nasal and a spirant. This would require, however, that affricates be clusters at some stage in the derivation.

The need for the study of the ways of utilising linguistic knowledge in speech is further confirmed by some surprising results obtained from direct phonological and grammatical tests. Earlier studies attempted to show that certain rules of the SPE phonology are not psychologically real as speakers fail to apply them to novel forms (nonsense words). Haber (1975) has shown that contrary to what might be expected speakers of English do very badly in tasks intended to test the productivity of the regular plural formation rule (the -(e)s ending), i.e. one that with good reason is generally assumed to be fully productive. It does not matter here whether the relevant mechanism is purely phonological, morphological or something else (the rule is transparent and could be formulated in surface terms). If tests fail to confirm the psychological reality of this simple rule, then most linguists would agree, I suppose, that there is something fundamentally wrong with the tests themselves; Kiparsky and Menn (1977, 64) ascribe it to "a "strangeness effect" which causes the subjects' performance to deteriorate relative to their normal speech" and are also (66-67) "skeptical about the ability of production tasks to show much of anything, at present, about the form of internalized linguistic knowledge, given the near-total obscurity surrounding the question of whether

and how this knowledge is used in speech".

As far as other areas of substantive evidence are concerned let me just mention two points: firstly evidence from an aphasiological study by Stark (1974) strongly suggests that the German velar nasal should be regarded as being derived from underlying /ng/, and this thus strengthens the case for an abstract interpretation of this problem vis-a-vis the stand taken by natural generative phonologists. Secondly, there is the case reported in Kiparsky and Menn (1977, 69-70) of an "invented language" which appears to exhibit two rules extrinsically ordered, which would indicate that the ordering of rules in itself cannot be difficult or impossible to learn as has been sometimes claimed. As Kiparsky and Menn point out, the charge that synchronic rule order mirrors diachronic developments cannot be made against speech invented by children.

The above discussion has not been meant to decry the importance of substantive evidence; conversely, in view of its potential significance I think it is necessary to stress that there is much in it which is arguable and which is itself in need of explanation and so can hardly be taken as definitive evidence for other theoretical concepts.

One final point that I would like to make is that the theoretical apparatus of abstract phonology is required to account for uncontroversially related, low phonetic details of pronunciation (see also Kiparsky 1975). Modifications, permutations, deletions and insertions of segments are well-known not only from abstract derivations but are also exceedingly common in accounts of rapid speech phenomena; thus, there is nothing basically new about abstract derivations that could not be found closer to the surface. Examples of the various modifications are well-known, and I would like to present a couple of examples from Polish where allegro rules introduce segments and contrasts totally absent from lento speech.<sup>2</sup> The phonetic inventory of Polish vowels contains six basic elements [i, ɨ, ɛ, a, ɔ, u], thus being again fairly regular typologically. Allegro forms introduce on the one hand a contrast of length which

(2) The examples are taken from Biedrzycki (1978) who interprets such data in terms of autonomous phonology and sets up phonemic distinctions for allegro styles which do not appear in lento styles.



does not appear in slow speech, e.g.: dała 'she gave' [da:] vs. da 'she will give' [da], stół 'table' [stu:] vs. stu 'of a hundred' [stu] corresponding to the lento forms [dawa - da] and [stuw - stu], respectively, and also several segments which are not known elsewhere, e.g.: in sp[ə:] czeństwo 'society' cz[ə:]m 'hi' - lento sp[ɔwɛ]czeństwo, cz[ɔwɛ]m; zapomni[ə:]m 'I forgot', chci[ə:]m 'I wanted' - lento zapomni[awɛ]m, chci[awɛ]m; cz[o:] 'one felt', ok[o:] 'one shod' - lento cz[uwɔ], ok[uwɔ]. The low level, optional rules which produce such forms are psychologically real and by producing new contrasts they seem to work like absolute neutralisation in reverse. If we were to postulate length contrast phonologically for Polish and then absolutely neutralise it, the abstractness sin would be committed; speakers of the language, however, seem to find nothing unusual about neutralising certain contrasts and introducing new ones when passing from lento to allegro styles. The force of these examples should not be overstated but they seem to show that there is nothing abnormal about rules merging and producing contrasts or about segments which appear at one level of representation but not at another.

The abstractness debate will no doubt continue both on language internal and external grounds. There remains much to do in both areas so that any final verdict at this stage would be premature.

#### References

- Biedrzycki, L. (1978): Fonologia angielskich i polskich rezonantów, Warszawa: Państwowe Wydawnictwo Naukowe.
- Dressler, W.U. (1977): rev. of Skousen, Substantive evidence in phonology, Lingua 42, 223-225.
- Fischer-Jørgensen, E. (1975): Trends in phonological theory, Copenhagen: Akademisk Forlag.
- Fromkin, V.A. (1971): "The non-anomalous nature of anomalous utterances", Lg. 47, 27-52.
- Goyvaerts, D.L. (1978): Aspects of the post-SPE phonology, Ghent: E. Story-Scientia.
- Haber, L.R. (1975): "The muzzy theory", in Papers from the 11th Regional Meeting CLS, R.E. Grossman et al. (eds.), 240-256, Chicago: Chicago Linguistic Society.
- Kenstowicz, M. and Ch. Kisseberth (1977): Topics in phonological theory, New York: Academic Press.
- Kiparsky, P. (1975): "What are phonological theories about?", in Testing linguistic hypotheses, D.Cohen and J.R. Wirth (eds.), 187-209, Washington, D.C.: Hemisphere Publishing Corporation.

Kiparsky, P. and L. Menn (1977): "On the acquisition of phonology", in Language, learning and thought, J.Macnamara (ed.), 47-78, New York: Academic Press.

Stark, J. (1974): "Aphasiological evidence for the abstract analysis of the German velar nasal", Wiener Ling. Gazette 7, 21-37.

Zwicky, A. (1975): "The strategy of generative phonology", Phonologica 72.

## THE PROBLEM OF PSYCHOLOGICAL REALITY IN THE PHONOLOGY OF PAPAGO

Kenneth Hale; M.I.T., Cambridge, Massachusetts, U.S.A.

This paper amounts to a clarification, for my own benefit more than anything else, of a certain linguistic principle involved in the evaluation of grammars. The principle, termed recoverability, is due to Jonathan Kaye (1975), and my conclusion about its nature and role in choosing among competing analyses is strongly influenced by a recent paper of Morris Halle's (1978). In examining the recoverability principle, I draw a comparison between two cases of alternative analyses - namely, the familiar problem of the Polynesian passive morphology, and a superficially similar problem in the phonology of Papago. In the course of the discussion, I will attempt to correct an error of reasoning which I made in an earlier treatment of the Polynesian case (Hale, 1973).

I begin with the Polynesian passive, using Maori to exemplify the classic situation. Some passives appear simply to involve straightforward suffixation of a vowel /-a/ - e.g., /patua/ beside active /patu/, /kitea/ beside /kite/, and so on. The majority, however, show a consonant between the root and a vocalic termination /-ia/ - e.g., /awhitia/ beside /awhi/, /hopukia/ beside /hopu/, /werohia/ beside /wero/, /inumia/ beside /inu/, etc. The consonant in these passives cannot be predicted from the surface form of the uninflected, or active, verb. There are at least two ways to analyze the consonantal passives. The 'phonological' analysis assigns the consonant to the stem, and the passive is formed by suffixing /-ia/ thereto. In addition, there is a rule deleting word-final consonants, thereby accounting for the fact that uninflected verbs, like all words in Maori, end in vowels. Thus, underlying /inum/ appears as [inu] if uninflected, but the deletion does not apply before the passive suffix, hence [inumia]. The 'morphological' analysis, by contrast, assigns the consonant to the suffix, thereby proliferating suffixal alternants (/ -tia, -kia, -hia, -mia.../), and each verb is assigned to a 'conjugation' according to the passive allomorph it selects. The assignment of the final consonant to forms like /inum/ is historically correct, and the deletion of these consonants in word-final position is a fact of the linguistic tradition leading to Polynesian. Nevertheless, I have attempted to argue (Hale, 1973) that the ahistorical morphological analysis is correct synchronically. If so, then what linguistic principle

dictates it? What motivated the change from the (historical) phonological analysis to the (ahistorical) morphological analysis? I suggested that the motivating factor was the canonical disparity, present in the phonological analysis, between the underlying morpheme structure of lexical items (allowing final consonants) and the surface syllabic canon of Polynesian (forbidding final consonants). The change to the morphological analysis eliminated this disparity.

Kaye has suggested another explanation. He defines a principle of 'phonological recoverability': "Recoverability concerns the degree of ambiguity manifested by a given surface form. The fewer the number of potential sources for the form, the greater its recoverability" (Kaye, 1975, 244-45). Phonological recoverability is valued in grammar, while 'phonological ambiguity', its converse, is devalued. Notice that the change in Polynesian completely eliminates phonological ambiguity - under the morphological analysis, the underlying form of a verb root is entirely recoverable from its surface form. Kaye proposes that phonological recoverability is the deciding factor in the Polynesian case.

This is a very promising suggestion. However, there is somewhat more texture to the problem which should be brought out in order to characterize the linguistic nature of the recoverability principle. As Halle (1978) points out, the Polynesian change did not really eliminate ambiguity. Rather, it shifted the ambiguity entirely to the morphology. The relation between uninflected and inflected forms remains ambiguous, since the derived form (the passive) is not predictable from the surface form of the active. Let us refer to this relation as 'morphological ambiguity' and to the converse relation as 'morphological recoverability'.

While phonological recoverability is logically distinct from morphological recoverability, it is not at all clear that the two principles are linguistically distinct. At least, I know of no convincing case in which phonological recoverability can be said to function autonomously in the evaluation of grammars. If the Polynesian change had consisted solely in the restructuring (i.e., in the realignment of the historic stem-final consonants onto the suffix), it would be possible, in principle, to argue that the change was motivated by phonological, rather than morphological, recoverability - since the former, but not the latter, would have been achieved. But the facts are different. The bulk of the evi-

dence which I adduced in favor of the morphological analysis consisted in observations to the effect that the conjugation system, assumed to have arisen through the restructuring of passive forms, was being regularized - a process which is complete in some Polynesian languages (e.g., Hawaiian, Tahitian) and merely well advanced in others (e.g., Maori, Samoan). I reasoned incorrectly that regularization implied restructuring. Surely regularization - i.e., the reduction of morphological ambiguity - could take place without restructuring. The evidence, therefore, does not directly support restructuring. Rather, it supports the view that recoverability is a genuine principle in the evaluation of grammars - assuming, as is reasonable, that change toward greater recoverability is in fact progressive. We cannot, on the basis of this evidence, at least, isolate phonological recoverability as linguistically distinct from morphological recoverability.

What, then, is left of the argument that the morphological analysis is correct in the case of the Polynesian passive? Before attempting to answer this question, let me introduce the Papago case (simplified in nonessential ways for the sake of space).

The points of interest can be illustrated by the third person singular possessed forms. These involve mere suffixation of [-j] to roots whose surface forms end in vowels - e.g., [mo'o] from [mo'o], [bahi] from [bahi], [gookij] from [gooki]. But when the root ends in a consonant in surface form, the suffix brings a vowel into view - e.g., [ñimaj] from [ñim], [hikaj] from [hik], [toonaj] from [toon], [ciñij] from [ciñ], [huucij] from [huuc], etc. Relevant historical events in the Piman tradition leading to Papago are the introduction of a palatalization rule, raising \*/t, d, n/ to [c, j, ñ] before high vowels, and the development of processes effecting the reduction or deletion of unstressed short vowels in certain environments - e.g., word-finally. Final short back vowels were deleted following any true consonant (i.e., nonlaryngeal), and final short \*i was deleted following coronals. While there is evidence that deletion was chronologically prior to palatalization, modern forms show the more natural nonbleeding order to have developed at some stage (e.g., [huuñ] from \*huunu). Since any of the five vowels of Piman (\*i, i, u, o, a/) could occur finally, deletion gave rise to ambiguity. This ambiguity is still present in the closely related Pima of Ónavas (in Sonora), where deletion

(but not palatalization) also exists - thus, for example, in Ónavas Pima, [hik] (from \*hiku) has the third singular possessed form [hikud], while [naak] (from \*naaka) has [naakad]. In Papago, however, the ambiguity has been entirely eliminated (in nouns, at least) through vocalic mergers. The deleting vowels merged as follows: (1) high vowels merged to /i/ following coronals; (2) back vowels not effected by (1) above merged to /a/. These mergers result in the circumstance that the vowel appearing in the suffixed form is recoverable from the quality of the surface final consonant in the uninflected base - if the consonant is a high coronal, the vowel is [i]; otherwise, the vowel is [a]. Thus, [hik] and [naak], ambiguous in Ónavas Pima, are recoverable in Papago.

Clearly, the elimination of ambiguity here is independent of any reanalysis of inflected forms which would associate the vowel with the suffix, rather than the stem, in forms like [hikaj] and [huucij] - the vocalic mergers in no way imply such a restructuring. It is entirely consistent with the facts to assume that modern Papago simply continues synchronically the historic deletion and palatalization rules (in nonbleeding order) and that the only restructuring consists in the vocalic mergers. While a restructuring to the morphological analysis would have achieved instantaneous phonological recoverability (in this area of Papago phonology, at least), there is no evidence suggesting that the change actually happened. Morphological ambiguity is the same under either analysis - namely, zero ambiguity.

Now let us consider the Polynesian and Papago cases together. What arguments can be constructed to choose an analysis in each instance? I think that the outcome will differ in the two cases, and, moreover, that the issue will turn on 'internal' arguments (cf. Kenstowicz, 1978) of a rather traditional sort.

I will assume, since I have no evidence to the contrary, that phonological recoverability is not distinct from morphological recoverability. Instead, there is a unitary principle of (morpho-phonological) recoverability according to which the value of a grammar increases as the amount of ambiguity (in relating base and derived forms) decreases.

In the Polynesian case, the phonological and morphological analyses are equal in terms of recoverability. This equality might be formalized, for example, by designing an evaluation metric

according to which the diacritic use of a phonological segment has the same cost as does an allomorph whose distribution is not predictable from surface phonology. Clearly, then, recoverability cannot be used to decide the issue here. From this fresh starting point, we see that there is no additional cost whatsoever associated with the morphological analysis. But there is an additional cost associated with the phonological analysis - namely, the deletion rule and, assuming it to be an extra cost, the canonical disparity between underlying morpheme structure and the Polynesian syllabic canon. Given these considerations, it seems to me that the rational choice here is the morphological analysis.

In the Papago case, likewise, recoverability fails to decide the issue. Here, however, there is nothing to recommend the morphological analysis. Its choice would not eliminate the necessity for the deletion and palatalization rules, since these are independently motivated - the morphological process of perfective truncation, among other processes, exposes medial vowels to the effect of the deletion rule, and a well motivated prevocalic vowel deletion rule exposes coronals to the palatalizing effect of suffix-initial /i/. Moreover, under the morphological analysis, we must distinguish at least two types of suffixes - one having a single alternant, continuing original Piman vowel-initial (e.g., the causative-benefactive formative [-id], from \*-ida), and another, continuing original consonant-initials and exhibiting synchronically three underlying forms distributed in accordance with an allomorphy rule (e.g., the modern forms deriving from Piman \*-di, [-j] after vowels, [-ij] after high coronals, [-aj] elsewhere). This second type of suffix, and the allomorphy rule associated with it, are entirely a product of the morphological analysis. There is no comparable cost associated with the phonological analysis. It does not, as it would in the Polynesian case, involve a canonical disparity, since underlying morpheme structure in the phonological analysis of Papago simply corresponds to the least marked of the rich variety of syllabic patterns admitted by the Papago canon. All things considered, the phonological solution here costs no more than what is necessary in a descriptive adequate account - it is, therefore, the rational choice in this instance.

I conclude from this discussion that recoverability is a genuine principle in the evaluation of grammars. Properly construed,

however, it enters into linguistic argumentation in much the same way as do traditional cost-accounting arguments which evaluate competing analyses in terms of relative parsimony. If this is correct, then it is not surprising that recoverability may fail to decide between alternative solutions - the alternatives may, as in the two cases examined here, be equal in terms of recoverability. Beyond recoverability, there are other principles which are relevant to the evaluation of grammars, including parsimony. I very much doubt that any of these principles can be said to carry greater psychological weight than others, i.e., to be more 'real' psychologically. Our task as students of language, it seems to me, is to determine which principles are justifiable linguistically - those principles will also be justifiable psychologically, given the subject matter of linguistic science. Of course, it is legitimate in making this determination to use evidence of all sorts, and some may prove to be more helpful than others.

#### References

- Hale, Kenneth (1973): "Deep-surface canonical disparities in relation to analysis and change", in Current Issues in Linguistics, Volume Eleven, T.A. Sebeok (ed.), 401-458, The Hague: Mouton.
- Halle, Morris (1978): "Formal versus functional considerations in phonology", manuscript, to appear.
- Kaye, Jonathan (1975): "A functional explanation for rule ordering in phonology", in Papers from the Parasession on Functionalism, R.E. Grossman et al. (eds.), 244-252, Chicago: Chicago Linguistic Society.
- Kenstowicz, M. (1978): "Functional explanations in generative phonology", manuscript, to appear.

## PSYCHOLOGICAL REALITY AND THE CONCEPT OF PHONOLOGICAL RULE

Per Linell, Dept. of Linguistics, Univ. of Uppsala, Sweden

1. Phonology is concerned with the sound patterns of various languages. In each language we use different sounds according to different rules, and the task of phonology is to define these rules. Thus, phonology is language-specific phonetics.

2. However, the usual phonological practice of most contemporary scholars in the field does not fit this description exactly. For example, in orthodox generative phonology many "low-level" language-specific phonetic regularities are not seriously considered, while many regularities which should actually belong to either lexicon or morphology are erroneously treated within phonology.

Though phonology should be concerned with speech and though speech is behavior, linguists have not studied it as behavior. Rather (some aspects of) the products of behavior have been studied in abstracto, i.e. idealized phonetic strings (words) and their interrelations have been analyzed without regard to how they are actually processed in speech production and perception and acquired by children, etc. Normally, the analysis is also crucially dependent on some kind of graphic representation. On this basis, the phonologist sets up a model of representations and rules which express connections between various idealized linguistic expressions and between properties of such expressions at various levels.

3. The problem of psychological reality in phonology concerns the relations between the representations and rules of the phonological model and the speaker-hearer's ways of storing and processing information about the structures of strings of phonetic behavior (their construction, pronunciation, recognition) and their interrelations.

4. The claims for psychological reality can be quite different in scope and content, ranging from those who assume an almost isomorphic relation between representations and rules in the phonological model and actually stored information and actual processes in speech performance, to those who see the relations as extremely indirect (the claims being therefore empirically empty). As for syntax, Fodor et al. (1974) are inclined to conclude that only the

(analysis of the) output of a standard GTG is psychologically valid. No doubt the same is true of an orthodox generative-phonological (OGPh) model (where, in practice, outputs are classical phonemic representations!). Underlying systems and derivations have no psychological reality or can be psychologically relevant only very indirectly.

5. Entities which are claimed to be psychologically valid should have plausible interpretations within (or at least be compatible with) a theory of meaningful linguistic behavior (speech). If we concentrate on phonology, i.e. on the phonetic aspects (aspects having to do with sound structure itself), what are the main problems that such a theory should be capable of solving? Perhaps the following should be mentioned:

- 5 a) How can we explain the fact that, although manifestations vary, there are many features that recur in the various manifestations of what speakers (of the same dialect) recognize as the same word form? I would propose that there is one common phonetic plan that defines the linguistic (phonological) identity of the word, a plan which specifies the linguistically relevant properties that speakers aim at realizing and which listeners tend to reinterpret into what they hear. (This is, I believe, the proper interpretation of the concept of "phonological form".)
- 5 b) How is it possible to construct phonetic plans for new forms that do not already exist as memory-stored forms? I assume that speakers may perform morphological operations which use memory-stored information to produce new phonetic plans as outputs. (These operations are naturally subordinated to the major (semantic, syntactic) intentions of the speaker's utterance construction.)
- 5 c) What is the nature of the memory-stored information used by morphological and syntactic operations?
- 5 d) How can we explain the language-specific variation in the possibilities of actually pronouncing and perceiving utterances, i.e. in the execution of utterance plans? (I assume that the phonetic aspects of an utterance plan would include at least the phonetic plans of the constituent words and a

prosodic plan of the utterance). To explain all the language-specific details of a particular utterance token, we would have to assume the existence of a fully specified articulatory plan that accounts for all the features that cannot be automatically ascribed to inherent properties of the speech apparatus. (Thus, note that the terms "phonetic plan" and "fully specified articulatory plan" are not synonymous.)

6. I have argued elsewhere (e.g. Linell 1979) that underlying morpheme-invariant forms and OGPh type derivations cannot be fruitfully incorporated into a plausible theory of meaningful phonetic behavior. Instead, there is some evidence that

- 6 a) phonetic plans (cf. 5 a) may be characterizable in terms of phonemic forms (general conditions on such forms may be stated in terms of "phonotactic rules").
- 6 b) some such phonetic plans are stored as lexical forms (stems, base forms, and some phrases) (cf. 5 b).
- 6 c) morphological operations take such memory-stored forms as inputs and produce new phonetic plans as outputs. If morphological operations are analytically split up into components, the components may correspond to morphophonological rules proper, and the whole operation will have a certain similarity to the abstract part of an OGPh derivation (except that the inputs are concrete phonetic forms rather than morphophonemic forms) (cf. (1) below).
- 6 d) the language-specific variations in normal, careful speech vs. sharpened (formal, expressive) speech and informal, casual ("fast", reduced) speech can be characterized in terms of phonological rules proper. Thus, fully specified articulatory plans may be derivable from the word-form-invariant phonetic plans (cf. 5 d).

7. In this paper I will discuss the proper interpretation of terms like rule, condition, operation, and process in phonology within a theory of the kind envisioned in §5.

Often, the discussion of the psychological reality of phonological rules is confused by the fact that several quite different concepts seem to be mixed up in most treatments.

- a) One is the (normal) interpretation of rule in the social sciences, i.e. as norm (or sometimes merely regularity) of behavior.
- b) Another one is the notion of mathematical rule, a mapping (or an instruction for the mapping) of one formally defined string of symbols onto another one.
- c) Since rules of type (b) are often described (talked about) as processes, i.e. changes of something into something else, it is sometimes tempting to interpret rules as performance processes.

The situation is further complicated in that empirically quite different sorts of regularities have often been regarded simply as "phonological rules". Thus, the putative similarities between morphophonological rules within a morphological operation like (1) and the "fast speech" rules relating different pronunciations of one and the same expression as in (2) are only superficial (and formal).

- (1) formation of noun from nonsense adjective according to the obscene-obscenity pattern:

Operand:	/rijs/
Morpholexical rule:	/rijs+it/
Trisyllabic laxing:	/risit/
Vowel shift:	/resit/

- (2) (from Donegan and Stampe, 1978) /plæntit/ plant it

Regressive nasalization:	pIæt̩t
Flapping:	pIæɾt
Progressive nasalization:	pIæ̃ɾ̃t

8. The basic concept of rule should be (7 a). Speech is a stream of phonetic behavior or phonetic events (that produce certain effects). What distinguishes speech from "mere vocalizations"

is the fact that the behavior must fulfil certain conditions of syntactic and phonological nature (and both speakers and listeners "know" this). In our model rules specify these conditions. Although behavior and actions are inherently processual, they can be looked upon either from the point of view of the processes themselves or as behavioral products. The latter is especially motivated as regards actions which are intended to produce certain effects. Thus, the act of pronouncing plant it in a certain, casual way [pɪ̃æ̃ɪ̃t] may be analyzed as follows: The speaker must construct a certain phonetic plan that corresponds to his communicative intentions, i.e. plant it rather than, e.g., plan it. This construction is thus subject to certain rules or conditions, which may be construed either as conditions on the behavioral operation (construction process) or on its effect (the resulting phonetic plan). The plan is then executed (realised, pronounced) in a certain way ([pɪ̃æ̃ɪ̃t] rather than [plæ̃ntɪ̃t]); the specifics of this pronunciation may be characterized as conditions on (rules for) either the pronunciation as a process or the pronunciation (or, rather, the fully specified articulatory plan) as product.

9. Note that rules concern properties of the intended behavioral products ("surface forms") (not some mystical morpheme-invariants). What these properties are must largely be determined by linguistic analysis. Thus, we cannot dispose of the traditionally linguistic (structural) analysis of language products (§2), although I would argue that (provided we are interested in psychological reality) this analysis must concern the products in relation to what we know about their production, perception, and acquisition (which means that observations of actual performance under normal and experimental conditions, slips of the tongue and the ear, child language, etc., will be of vital importance).

10. Obviously, rules as generative systems (in e.g. the OGPh fashion) need not have anything to do with conditions on actual (or potential) behavior. Indeed, the idea that behavior could be governed by generative systems seems very naive. (The various figures of figure-skating could no doubt be specified by a generative theory of figures, but who believes that the skater's behavior is produced by means of processes corresponding to such generative rules?) Thus rules are not acts or processes, but conditions on behavioral acts or on their products.

11. Behavior can be talked about at several levels of abstraction. When we talk about the morphological operations of constructing e.g. /resɪtɪ/ from /rɪ̃s/ (cf. (1)) or /fɔksɪz/ from /fɔks/ (pluralization), we are not necessarily modelling the actual behavioral process. The only thing we can say is that there is evidence that speakers can (sometimes) form "correct" ity-nouns from nonsense adjectives, that they can form plurals of English nouns, and that the respective operations are subject to certain linguistically defined conditions. That is, we can assume that speakers actually carry out morphological operations and other linguistic actions (and our models specify the linguistic content of the actions), but we cannot speculate on how these operations are neuro-physiologically implemented. Operations and actual processes lie at two different levels of description and must not be identified. Operations are defined by their intended effects, and it is conceivable that there are many ways for the neural mechanisms to achieve the goals.

It follows that rules must not be equated with behavioral processes. Not even in casual speech phonology are we entitled to conclude that rules correspond to processes. After all, conventional phonological rules state nothing but regular correspondences between idealized representations of the same or related pronunciations. (Note that I am not using 'rule' and 'process' in the way they are used in Stampean "natural phonology".)

12. I started by defining phonology as language-specific phonetics, and later I characterized rules as norms. However, this means that the phonology of a specific language would not describe or explain all the details of actual pronunciations in that language, since not all facts are conventional; some follow from biologically determined limitations. (In casual speech phonology, most regularities are language-specific variants of otherwise universal phonetic tendencies.) This is a reasonable definition of phonology, since it confines phonology to those features that must be learnt. However, we could alternatively generalize 'rules' to cover all regularities, whether conventional or biologically determined. Such a conception seems to be accepted in Stampean phonology. Thus, e.g., children's incompetence rules (i.e. Stampe's inherited processes) are clearly not social conventions. But even

such rules remain correspondence formulas; the actual phonetic processes are probably more of general continuous adjustments along scales.

13. The analysis of concepts like "psychological reality", "rule" versus "process" and "operation", etc. is necessary if the relation of phonology to phonetics is to be properly understood.

Acknowledgement

I want to thank Sven Öhman for much inspiration through the years.

References

- Donegan, P.J. and D. Stampe (1978): "The study of natural phonology", forthcoming in Current approaches to phonological theory, D. Dinnsen (ed.), Bloomington: Indiana University Press.
- Fodor, J.A., T. Bever and M. Garrett (1974): The Psychology of Language, New York etc.: McGraw-Hill.
- Linell, P. (1979) (forthcoming): Psychological Reality in Phonology, Cambridge University Press.
- Steinberg, D. (1975): "Chomsky: From formalism to realism and psychological invalidity", Glossa 9, 218-252.



## EMPIRICAL INTERPRETATIONS OF PSYCHOLOGICAL REALITY

Royal Skousen, University of Texas, Austin, Texas, USA

In this paper I will discuss three requirements for a theory of language. These requirements are (1) inducibility, (2) generality, and (3) testability.

The first requirement, that of inducibility, is that linguistic descriptions must be directly derivable from the data that speakers are actually confronted with in learning the language. A linguistic description thus implies (1) a description of the relevant data and (2) a set of rules by which the linguistic description is derivable from the data. We refer to this set of rules as the rules of induction.

In order to understand this requirement of inducibility, let us consider some common violations of this requirement. For example, the order and frequency with which the data is presented to the speaker may be significant in determining the proper description of the data or in explaining how the language may change over time, so that if such information is ignored, the subsequent description may be untestable. Consider, for instance, Chomsky and Halle's statement in The Sound Pattern of English (p. 332) that "it is no doubt the case that the linguistic forms that justify our postulation of the Vowel Shift Rule in contemporary English are, in general, available to the child [?] only at a fairly late stage in his language acquisition, since in large measure these belong to a more learned stratum of vocabulary." Of course, there is no way that Chomsky and Halle's description itself can be empirically tested, since their description is based on data that, as they themselves admit, is unrepresentative of the data that children are confronted with in learning English. Children learn to speak long before they learn words as infrequent as profanity, comparative, gratitude, serenity, appellative, plenitude, divinity, derivative, conciliate, and so forth (SPE, p. 50).

Another common violation of inducibility occurs when a non-existent ordering is imposed on the data. A common method of explicating linguistic data is to first offer that data which provides direct evidence for some rule and then treat the exceptions to the rule afterwards - by adding additional rules perhaps, but without changing the original rule. Consider, for instance, Chomsky and Halle's treatment of Kasem singular and plural forms in SPE (pp. 358-364). They first give us "regular" forms like bakada and bakadi

(singular and plural for 'boy') as evidence that the singular ending is a and the plural ending is i. Then they give us the surface exceptions to this "regularity" (e.g. kambia/kambi 'cooking pot', pia/pi 'yam', buga/bwi 'river', diga/di 'room', laŋa/lə 'song', naga/nə 'leg', pia/pə 'sheep', and so on). Chomsky and Halle try to explain these forms without abandoning their original "regularity", but their explanation depends crucially upon the order of presentation of these "irregular" forms. For instance, they first argue that a plural form like kambi can be considered "regular" (that is, as /kambi+i/) if there is a phonological rule of truncation that will reduce ii to i. Having thus established that the "regular" endings are a and i and that there is a rule of vowel truncation, then a singular form like pia 'sheep' can be interpreted as /pia+a/: "Since the grammar already [!] contains the Vowel Truncation Rule, [pia] can also be derived from an underlying [piaa]." From an acquisitional point of view, Chomsky and Halle are assuming that the speaker takes care of the "regular" cases first and then the "irregular" case kambi before tackling the "irregular" case pia 'sheep' (sg.)' (which, incidentally, is "regular" on the surface). Finally, Chomsky and Halle posit a rule of metathesis for Kasem, again assuming that all rules previously posited will be maintained. The rules which Chomsky and Halle present depend upon their artificial ordering of the data. But the data is not ordered in this way for the child learning Kasem, nor does the child know in advance which of these forms are "regular" and which ones are "irregular" or "exceptional". If such a characterization of these forms is correct, then the child must discover it from random data.

Another violation is to ignore some of the data, especially those forms which the linguist knows are "incorrect": slips of the tongue, false starts, analogical creation, stuttering, dialectal variants, and so on. Yet the child does not know in advance which of the forms in the data are errors. If a child hears another child using the form goed for the past tense of go, we do not delete this from the child's data. We keep it in the data, but try to explain why the child will eventually identify goed as an incorrect past tense form. Nor should we even delete examples of stuttering from the data, since speakers can learn to imitate stuttering. Speakers also learn how to show that they have made a false start. For instance, speakers of English may use uh (but not /i/) to indicate a false start. Nor do speakers ignore dialectal variants - they learn them, even though they may not use them.

Finally, linguistic descriptions cannot be based on non-existent data. Although speakers can learn that certain items do not occur in the data, this knowledge cannot be derived from knowing in advance that these items do not occur. The determination, for example, of syntactic descriptions cannot depend upon knowing which non-occurring sentences are ungrammatical and which ones are grammatical.

The second requirement is that of generality: The rules of induction are independent of any given set of linguistic data and independent of any given regularity found in linguistic behavior. In other words, the rules of induction are universal and not taxonomic or ad-hoc. Only in this way can the explanatory goal of linguistic theory be achieved.

An excellent example of a universal rule of induction is found in Jakobson's Child Language, Aphasia, and Phonological Universals in which Jakobson proposes that "the sequence of stages of phonemic systems" found in such diverse areas as aphasia and the acquisition of languages "obeys the principle of maximal contrast and proceeds from the simple and undifferentiated to the stratified and differentiated" (p. 68). Of course, there are problems with some of Jakobson's specific claims about language acquisition, aphasia, and the phonemic systems of the languages of the world. Nonetheless, the significant contribution that Jakobson makes is that he proposes a conceptually simple and universal principle in order to explain a diversity of linguistic behavior.

In accordance with Jakobson's general principle, let us consider a principle of maximizing acoustic differences and see how it might explain the instability of certain sounds in the languages of the world. Take, for example, the case of the phoneme ü. In comparison to the phonemes i and u, ü is unstable and relatively infrequent. Children trying to learn a language that has the phoneme ü generally replace it with i or u. Historically, languages with ü frequently lose it in favor of either i or u. In the languages of the world we find phonemic systems with i-u and i-ü-u, but i-ü is relatively rare, and ü-u, as far as I am aware, is non-existent. And when i-ü does occur, it is unstable and is usually replaced historically by the more stable phonemic systems i-u and i-ü-u. Finally, when an adult speaker of an ü-less language attempts to pronounce ü, it will be pronounced as i, u, or perhaps the diphthongal iu. Now Chomsky and Halle "account for" this linguistic behavior by means of a taxonomic marking convention which simply

recapitulates the linguistic behavior formalistically (SPE, p. 405):

$$[u \text{ round}] \rightarrow [\alpha \text{ round}] / \left[ \begin{array}{c} \text{aback} \\ \text{-low} \end{array} \right]$$

But a principle of maximizing acoustic differences could be used to explain this behavior. The motivation for this principle is that small acoustic differences are difficult to perceive and produce, thus shifts will occur in the direction of increasing acoustic differences. If we consider the first three formants of the vowels i, ü, and u, the maximal distinction occurs between i and u and thus the intermediate ü may be replaced by the phonetically similar i or u.

The important point in using a general principle such as this one is that it can account for the linguistic behavior of other sounds besides ü. For instance, the interdental fricatives θ and ð are also unstable and infrequent and tend to be replaced by phonetically similar sounds such as the dental fricatives s and z, the labiodental fricatives f and v, or the dental stops t and d. On the other hand, Chomsky and Halle's approach leads them to postulate a completely different marking convention in order to handle the instability of the interdental fricatives (SPE, p. 407):

$$[u \text{ strid}] \rightarrow [\alpha \text{ strid}] / \left[ \begin{array}{c} \alpha \text{del rel} \\ \left\{ \begin{array}{l} [+ant] \\ [+cor] \end{array} \right\} \end{array} \right]$$

Such taxonomic rules do not explain anything; they merely formalize observed regularities. The observation of regularities is, of course, critical to the construction of a theory, but observed regularities do not make theories. Instead, regularities demand explanation in terms of general principles.

The third requirement for a theory of language is that it must be testable: A theory must have an empirical interpretation. Let us assume that we have some linguistic data for a particular language and that we apply certain rules of induction to the data and derive a description of the data. The question of utmost importance is: How can we discover if the proposed rules of induction are correct? In other words, how can we determine if the linguistic description really represents what the speaker has learned? It is not enough to simply declare that the description is psychologically real. The linguistic data is available for observation, but we cannot observe

the rules of induction that speakers are using to learn the language nor can we observe the derived linguistic descriptions. But we can observe subsequent linguistic behavior. So in order to test the rules of induction and the derived linguistic description, we need a mapping between the linguistic description and linguistic behavior. This mapping is the empirical interpretation. A theory is tested by its ability to predict the nature of linguistic behavior. Thus a theory is composed of two parts: (1) the rules of induction and (2) the empirical interpretation of descriptions. A theory without an empirical interpretation is not really a theory because it is not testable. Most so-called "theories" of language are actually rules of induction - that is, systematical methods for describing linguistic data (or deriving linguistic descriptions). Theory construction must also include the interpretation of descriptions. The empirical interpretation will predict how speakers would use the linguistic description. By comparing the predicted behavior with actual behavior we can test our theory. If a theory has an empirical interpretation (that is, if the theory is falsifiable), then we may ask if there is any evidence in favor of this theory over alternative theories and if there is any evidence against this theory. If the theory fails in some respect to correctly predict actual linguistic behavior, then the fault may lie in the rules of induction or the empirical interpretation, presuming that the linguistic data is accurately represented.

A good example of an empirical interpretation of a linguistic construct is found within those phonological theories that treat the phoneme as a psychological unit. Consider, for instance, the following possible empirical interpretations of the phoneme:

(1) Naive spellings (especially the spellings of children learning how to read and write) are based on phonemic representations. On the basis of this empirical interpretation, Read (1975, 29-78) argues that invented spellings like CHRIE for try, JRAGIN for dragon, NUBRS for numbers, LITL for little, and LADR for letter give evidence that the children's phonemic representations for these words are /čraj/, /jɾəgən/, /nɔ̃brz/, /lɪtl/, and /lɛdr/, rather than the more common phonemic representations /traj/, /drəgən/, /nɔ̃mbərz/, /lɪtəl/, and /lɛtər/. (These latter forms have undoubtedly been influenced by the standard orthography.) Similarly, Sapir argued (1968, 54-58) that his informants' naive spellings were also representative of their phonemic representa-

tions.

(2) Slips of the tongue are based on phonemic representations. For instance, Fromkin (1971, 33) argues that since slips of the tongue never split apart the affricates [tʃ] and [dʒ] in English, these affricates should be interpreted as single phonemes, /č/ and /ǰ/, rather than as a sequence of phonemes, /tʃ/ and /dʒ/. In contrast, actual phonemic sequences like [spr], [pɪ], [kr], [bɪ], and [fr] are frequently split apart. This difference in linguistic behavior is explained if we assume that this empirical interpretation is correct. Similarly, Stampe (1973, 35) argues that there are no archiphonemes in English because of the occurrence of [hwɪpsr̩] rather than [hwɪbsr̩] as a slip of the tongue for the word whisper. The psychological (or phonemic) representation of whisper is, say, /hwɪspr/ rather than /hwɪsbr/ or /hwɪsBr/, where B stands for a labial stop unspecified for voicing (that is, an archiphoneme). The reason then that the slip of the tongue is [hwɪpsr̩] is that slips of the tongue switch the order of phonemes, and the metathesis in this example shows that the real phonemic representation contains a voiceless, bilabial stop.

(3) Linguistic games are based on phonemic representations. This empirical interpretation serves as the basis of Sherzer's (1970) analysis of the Cuna language. The games that speakers play are characterized as simple operations on strings of phonemes, although one speaker's phonemes may be more "abstract" than another's. The problem of the English affricates can also be studied by means of linguistic games. Those "speakers" of Pig Latin who move only the first consonant of an initial consonant cluster (e.g. spin is [p<sup>h</sup>ɪnsɛj]) always move the complete affricate (e.g. chin is [ɪnčɛj] and never [sɪntɛj]), thus indicating once more that [tʃ] is to be interpreted as a single phoneme, /č/, rather than as /tʃ/.

Now let us suppose we have some rules of induction for the determination of phonemic representations and that these rules lead to the interpretation that the English affricates should be sequences of phonemes, as /tʃ/ and /dʒ/. Without any empirical interpretation of phonemic representations, there would be no way to test this description of English or the rules of induction which are used to derive this description. In order to test our theory, we must determine some empirical interpretation for our phonemic representations. If we accept these three interpretations (namely, that naive spellings, slips of the tongue, and linguistic games

are based on phonemic representations), then we can test this description of the English affricates and any set of inductive rules that would lead to such a description. We have already seen that the evidence from linguistic games and slips of the tongue imply that the affricates are unitary. In fact, children's spellings also support this conclusion, since there is no evidence for invented spellings of the form TSH for the affricate /č/ (e.g. chin is not spelled as TSHIN). In this case, all three empirical interpretations argue against the phonemic representations /tʃ/ and /dʒ/. These interpretations support each other, which is what we should expect if all three of these interpretations are correct. Now it may be that these empirical interpretations are, in fact, incorrect, but we should not reject them simply because we desire, above all else, to maintain our description of the English affricates (as /tʃ/ and /dʒ/) and the rules of induction that derive them. And even if these empirical interpretations are not correct, this does not relieve the phonologist of the responsibility to provide some empirical interpretation for his phonemic representations. In order for his theory to be testable, the linguist must determine what will count as evidence for his description and what will count against it. If the linguist can think of nothing that will disprove his theory, then he does not have a theory.

One important empirical interpretation that should hold for any theory is the principle of homogeneity: If the rules of induction do not distinguish between A and B in the linguistic description, then the behavior of A and B should be the same. Thus the rules of induction can be shown to be wrong if, in fact, A and B behave differently. The principle of homogeneity requires linguistic theory to predict linguistic behavior accurately. If a theory fails to predict an observed difference in linguistic behavior, then the theory must be revised.

A well-known case where this general principle of empirical interpretation has been used is in Kiparsky's paper "How Abstract is Phonology?". Kiparsky argued (pp. 24-25) that "contextual neutralizations are reversible, stable, and productive, whereas the alleged absolute neutralizations are irreversible, unstable, and unproductive." Now the standard generative phonology of that time did not distinguish between contextual and absolute neutralization; both were equally possible. Since linguistic behavior does distinguish between these two categories, the theory must be wrong.

Kiparsky therefore argued that the theory must include an alternation condition, which would either forbid absolute neutralization or at least make it highly improbable. In this way the linguistic theory could predict the non-homogeneous linguistic behavior.

This example suggests that the principle of homogeneity can be used to discover what sorts of information a linguistic description should have in order to predict differences in linguistic behavior. In fact, without the goal of predicting linguistic behavior, there would be no motivation for discovering the psychologically real linguistic descriptions.

#### References

- Berko, Jean (1958): "The child's learning of English morphology", Word 14, 150-177.
- Chomsky, Noam and Morris Halle (1968): The Sound Pattern of English, New York: Harper and Row.
- Fromkin, Victoria (1971): "The non-anomalous nature of anomalous utterances", Language 47, 27-52.
- Jakobson, Roman (1972): Child Language, Aphasia, and Phonological Universals, The Hague: Mouton.
- Kiparsky, Paul (1973): "How abstract is phonology?", Three dimensions of linguistic theory, Osamu Fujimura (ed.), 5-56. Tokyo Institute for Advanced Studies of Language, Tokyo.
- Read, Charles (1975): Children's categorization of speech sounds in English, National Council of Teachers of English, Urbana.
- Sapir, Edward (1968): "The psychological reality of phonemes", Selected Writings of Edward Sapir, David G. Mandelbaum (ed.), 46-60, Berkeley: University of California Press.
- Sherzer, Joel (1970): "Talking backwards in Cuna: The sociological reality of phonological descriptions", Southwestern Journal of Anthropology 26, 343-353.
- Stampe, David (1973): A dissertation on natural phonology, PhD dissertation, University of Chicago.

## ACQUISITION OF THE PHONOLOGICAL SYSTEM OF THE MOTHER TONGUE

## Summary of Moderator's Introduction

Charles A. Ferguson, Department of Linguistics, Stanford University, Stanford, CA 94305, U.S.A

The papers are fairly representative of the range of studies of phonological development now being undertaken. The period 1968-78 was one of greatly increased research in child phonology, in large part stimulated by the English translation of Kindersprache (Jakobson 1968) and the publication of experiments on speech-sound discrimination in infants (Eimas et al. 1971). A recent conference attempted to review and synthesize this research (Yeni-Komshian et al., in press).

The three areas of greatest current research effort are: neo-nate discrimination, the transition from babbling to speech (first two years of life) and the development of phonological organization (age 2-4 yrs.). The first and third are represented here directly by Kuhl and Menn, respectively, and all three are alluded to in the various papers. The phonetic/phonological development of older children is represented here by Gilbert and Hawkins, and the Hawkins paper represents the expanding field of the development of prosodic and temporal characteristics of speech.

The papers are also representative of new trends in research orientation. Earlier emphasis on innate structures and processes led to concern with (a) universal orders of acquisition of phonemes, features, and phonological oppositions, and (b) the identification of feature detectors roughly analogous to visual feature detectors. The new trend is toward emphasis on variation in the order and routes of development and on the effect of input on the child's development. The emphasis on variation is striking in Menn's paper, which classifies variation into seven types and relates these to possible developmental models, but it is evident also in Menyuk's paper, which notes that "universality is confounded by the particular data the child is confronted with." Similarly, Kuhl, who is concerned with species-wide predispositions and even predispositions shared with other species, examines the importance of "selective auditory exposure" and concludes that in infants' speech-sound category formation "their tendencies to attend to particular acoustic dimensions [are] modified by exposure to a particular language."

Another new trend is the reversal of earlier confidence that

adult models of speech processing and linguists' phonological theories are good bases for understanding child phonology. Current research tends to claim that the contribution may often go in the opposite direction, that developmental studies may help in understanding adult models and may offer a valuable corrective to phonological theory. The models offered by Menyuk and Menn, although different in approach, both illustrate this trend. Menyuk's "outside-in" model is deliberately different from linguistic segmentation and hierarchization and also suggests that current adult models may be inadequate even for adults. Menn's two-lexicon model with both non-automatic and automatic production processing has implications, not much explored by her here, for an adult model of phonology which would allow for more variation than most theories.

Knowledge of infant speech perception is increasing rapidly, as several research paradigms are followed (Morse 1974, Kuhl in press). Knowledge of "pre-linguistic" speech production is likewise increasing (cf. Dore, et al. 1976, Stark 1978, Carter 1978). Finally, both data-oriented and model-oriented studies of the development of phonological structure are increasing (e.g. Ferguson and Farwell 1975, Kiparsky and Menn 1977).

Unfortunately, however, the conceptual gap between the infant perception studies and the other studies seems to be widening. The former are perception-oriented and elaborately experimental, the latter are production-oriented and based on naturalistic observations, typically of a small number of subjects or even a single child. Ways must be found to study perception in pre-school children and to connect the neo-nate studies with studies of older children (cf. Strange and Broen, in press).

#### References

- Carter, A.L. (1978): "From sensori-motor vocalizations to words", in Action, gesture and symbol: the emergence of language, A. Lock (ed.), 309-349, London: Academic Press.
- Dore, J., M.B. Franklin, R.T. Miller and A.L. H. Ramer (1976): "Transitional phenomena in early language acquisition", J.Ch.Lang. 3,13-28.
- Eimas, P., E. Siqueland, P. Jusczyk and J. Vigorito (1971): "Speech perception in infants", Science 171,303-306.
- Ferguson, C.A. and C.B. Farwell (1975): "Words and sounds in early language acquisition", Lg 51,419-439.

- Jakobson, R. (1968): Child language, aphasia and phonological universals, The Hague: Mouton.
- Kiparsky, P. and L. Menn (1977): "On the acquisition of phonology", in Language learning and thought, J. Macnamara (ed.), New York: Academic Press.
- Kuhl, P.K. (in press): "The perception of speech in early infancy", in Speech and language: research and theory, N.J. Lass (ed.), New York: Academic Press.
- Morse, P.A. (1974): "Infant speech perception: a preliminary model and review of the literature", in Language perspectives-- acquisition, retardation, and intervention, R.L. Schiefelbusch and L.L. Lloyd (eds.), 19-53, Baltimore: University Park Press.
- Stark, R.E. (1978): "Features of infant sounds: the emergence of cooing", J.Ch.Lang. 5,379-390.
- Strange, W. and P.A. Broen (in press): "Perception and production of approximant consonants by three-year-olds: a first study", to appear in Yeni-Komshian et al. (eds.) Child Phonology: perception, production and deviation, New York: Academic Press.
- Yeni-Komshian, G., J.F. Kavanagh and C.A. Ferguson (eds.) (in press): Child phonology: perception, production and deviation, New York: Academic Press.

ON THE VOWEL AND ITS NATURE, BETWEEN EIGHTEEN MONTHS AND FIVE YEARS

John H. V. Gilbert, Phonetics Laboratory, Division of Audiology and Speech Sciences, University of British Columbia, Vancouver, Canada

### Introduction

For a number of years, we have been interested in the development of vowels in children between eighteen months and approximately five years of age (chronological age, CA), and have conducted a number of studies which have been directed toward questions in both production and perception. In this paper I shall review them, and some other studies which relate to the title, (Gilbert a-d).

Our original curiosity about the development of vowels was motivated largely by two factors; the first being that in 1967, there was very little information relating to this particular aspect of phonological acquisition; the second being that the classic paper of Peterson and Barney (1952) tantalizingly showed marked differences in vowel formant measure between children and adults without at any place in the paper stating how old the subjects were who constituted their sample. The Peterson and Barney data showing the differences between adult males and children are illustrated in an F1/F2 plot shown in Figure 1.

Our interest in the development of vowels then developed into

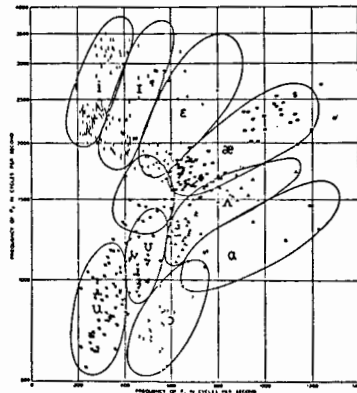


Figure 1. Frequency of second formant versus frequency of first formant for vowels spoken by men and children, which were classified unanimously by all listeners (Peterson and Barney, 1952).

three principle questions: the first was whether it was possible to

accurately trace the development of vowel sounds from around eighteen months to their adult values; in this we were superceded by the excellent work of Eguchi and Hirsch (1969) whose formant measures for vowels over time are shown in Figure 2.

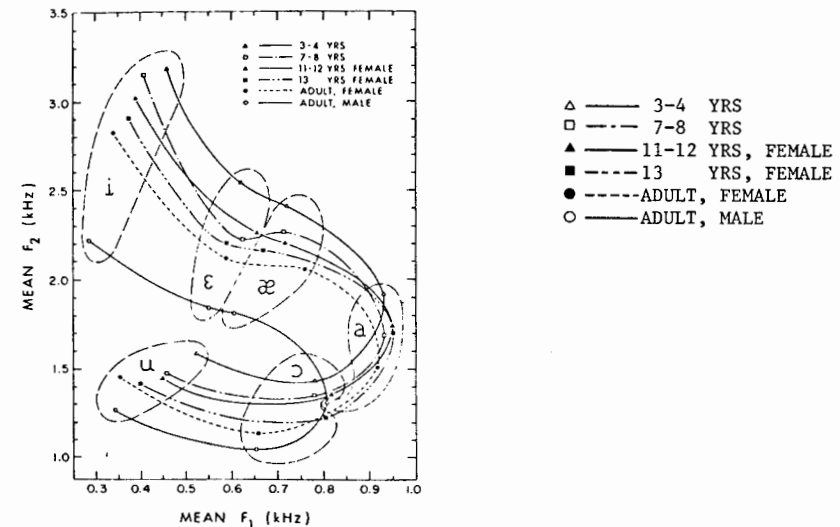


Figure 2. Mean formant frequencies for combined age groups as shown in the key. Each point represents the combination of Formant 1 and Formant 2 for each of the six vowels. The different symbols together with the lines that join them represent the different ages. The broken circles are drawn around all points for a given vowel. (Eguchi and Hirsh, 1969).

There were, however, subsidiary questions relating to the problem of the ontogeny of vowels, in particular, whether children of the same chronological age but at different stages of physiological development, would demonstrate differences in vowel formant frequencies because of their differences in growth. I will report this information later.

The second principle question related to formant measures of vowels produced by groups of children who were measurably different in their linguistic development, since a great deal of space in the phonological literature has been (and continues to be) devoted to a discussion of how and in what sequence consonant sounds emerge. We felt that children at different stages in the acquisition process might give us some information relating to this question, at least for vowels.



A third, and last principle question, concerned the manner in which vowel sounds are perceived by children when these sounds are produced both by themselves and adults. Since vowels are perhaps more easily acoustically measured than consonants in the output of children, and since there appears to be more listener agreement on their character, we considered this line of investigation was one worth following.

#### Studies of Vowels

Because of their clear separation in the vowel quadrilateral, our energies were directed chiefly to an examination of four vowels: /i/ as in "heed", /æ/ as in "had", /ɒ/ as in "hod", and /u/ as in "who'd", produced by both children and adults, usually in an h-d environment. Other studies reported in the literature have examined a wider set than this. The choice of these vowels, however, allowed us to compare our results with results from numerous studies conducted with adults, and in retrospect, to consider some issues, e.g. individual variation, as they apply to the emergence of vowels during acquisition. Bearing in mind the problems of holding "mechanical" (i.e. child) parameter constant, and the difficulties of minimizing measurement variation (see Kent, 1976, 1978 for details on acoustic analyses of children's vowels), we hoped to view the vowel system "settling down" across chronological age.

In an early paper, Okamura (1966) measured five vowels spoken by 475 Japanese children and demonstrated that the formant frequency construction of these vowels was quite different between children and adults. A copy of his centre formant frequency measurements is shown in Figure 3.

It will be seen that for all of these vowel sounds, the formant frequency measurements appear to plateau around seven years of age. When we came to compare our own data for four-year-old English speaking children with that of Okamura, we found a fair measure of agreement for formant two. Our measurements are shown in Table 1.

Interestingly, the use of duration of vowels in emerging phonology appears, at least in one report (Di Simoni, 1974), to follow this development trend; by age six, durational differences between vowels becoming stabilized in children's speech. This issue is, however, confused and the reader is referred to a comprehensive account of factors in Greenlee (1978).

As mentioned earlier, one of our interests was to determine

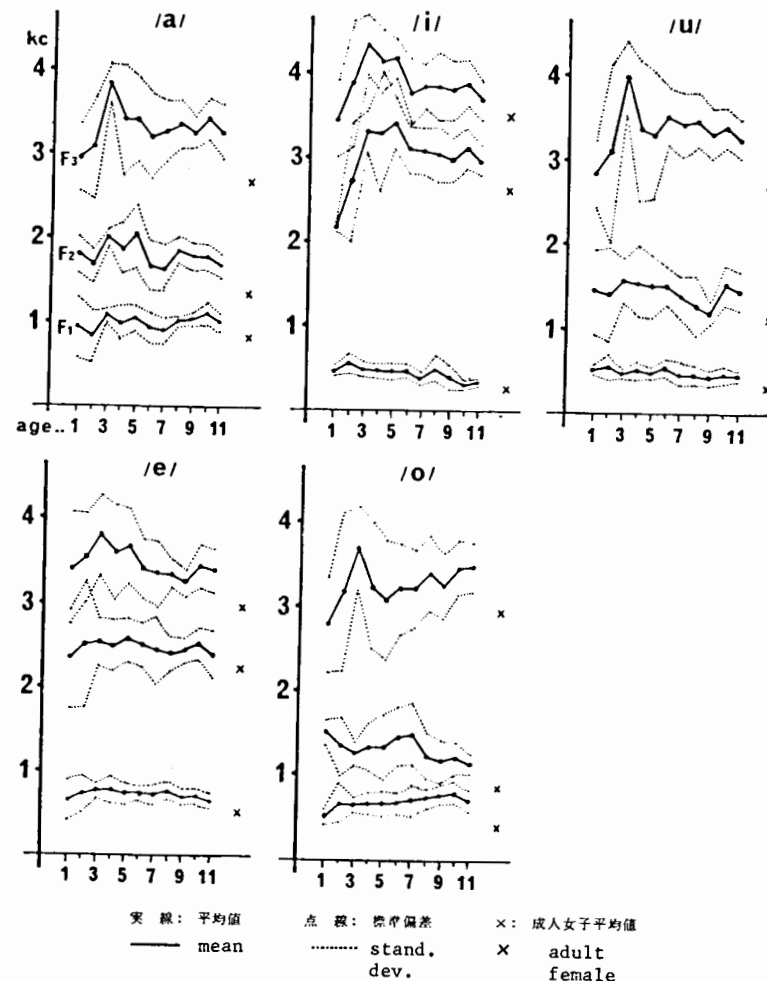


Figure 3. Formants and their standard deviations (Okamura, 1966)

wether differences in physiological age (whilst holding chronological age constant) would, in fact, change the acoustic characteristics of children's vowel sounds. It did not appear sensible to group children by CA for the purposes of examining vowel development if, in fact, their physiological ages were markedly dissimilar. The motivation for this observation was the assumption that a difference in physiological age would mean a difference in vocal

tract length and therefore a difference in characteristic vocal tract resonances.

Table 1. Vowel productions: means and standard deviations, in hertz for F1 and F2 measurements of control and experimental groups (Gilbert, 1970).

Vowel		F1		F2	
		Mean	Stand. dev.	Mean	Stand. dev.
/i/	Control	442	107	2510	99
	Experimental	555	149	2613	67
/æ/	Control	917	183	1710	251
	Experimental	859	130	1631	122
/a/	Control	693	112	1246	157
	Experimental	727	113	1216	299
/u/	Control	539	166	1255	202
	Experimental	533	115	1336	207

We found that both F1 and F2 naturally show a tendency to drop with an increase in chronological age from fourteen to eightyfour months, and that when subjects were reassigned to groups by Bone Age (BA) (Harrison, et al. 1964) groupings, (BA being the physiological measure which we used), the same pattern emerged. We found no statistical difference between Ca and BA on our formant measures; we thus concluded that grouping children by a measure of physiological maturity (rather than CA) does not in the final analysis alter results.

In retrospect, I am not sure that this was an appropriate conclusion to draw, based on the way in which we assigned children to BA groupings. I suspect that it would have been more appropriate to have taken both BA plus skeletal size, i.e. height and weight, and then compared them with children of similar CA and intelligence. We know from the work of Negus (1949) that the larynx develops most rapidly between 3;0 and 5;0 CA and then increases in size to maturity very slowly. This point should have been taken into account. I am still not convinced that we have solved the physiological age problem in our deliberations.

From a consideration of physiological age we then moved to a slightly different view of the process, that is, would children at different levels of linguistic development, but at the same CA exhibit any significant differences in vowel production. Given that children are the same height and weight our assumption would be

one of no difference, since we would expect that whatever emerged from the vocal tract would be of the same order, regardless of whether or not each child's linguistic abilities were different. We recorded children at 4;0 CA divided into groups on the basis of normal and late language usage. Although there were no differences between these groups in terms of mean formant two measurements, when we played tokens produced by the late language users to adult listeners for identification, the adults were definitely confused in their perception, a result which we had certainly not anticipated.

We interpreted this discrepancy as an indication that children who are at a less mature stage of linguistic development are doing "something" to the vowels which cannot be accounted for on an acoustic basis. A thorough examination of the acoustical similarities and dissimilarities between normally developing children and language delayed children is necessary before we can make any further judgements. It may well be that the dynamic acoustic information distributed over the temporal course of the syllable, is affecting listener judgement differently in each case.

The last question to which we addressed ourselves involved the perception of vowels. In 1967, Menyuk reported an experiment in which she showed that the phoneme boundaries for a set of vowels in consonantal context were the same for six children between 5;0 and 10;0 as they were for adult listeners. We found in our experiments that children at 4;0 have no difficulty in discriminating four broadly spaced vowel tokens spoken by themselves and by adults, when these are presented to them in an h-d context. We also found that, when children at this age are asked to produce vowel sounds in an h-d context (in response to the same vowels in h-d context spoken by adults and child speakers other than themselves), there is virtually no difference between F0 and F2 in their tokens, and the tokens of the speakers whom they are imitating.

Lieberman (1978) and his associates at Brown University have data which shows a gradual and consistent improvement in the children's productions of vowels of English from the early stages of babbling through to 3;0; an age at which the children are using meaningful sentences and conversing with the experimenters. Lieberman's data is very robust, and certainly corroborates our own notions about vowel development.

Conclusion

The question of the acoustical development of vowel sounds appears to be reasonably well answered by now. That is, one sees an increasing trend over the first six years towards the adult form in terms of fundamental frequency, F2 and F3. The plateauing between 6;0 and 8;0 is undoubtedly related to the fact that the vocal tract at this time is approaching its adult measurement. The question of the child's perception of vowel sounds appears more equivocal. Since perception will have to be accounted for by correct usage in production, we will need further experiments of the kind recently reported by Greenlee (1978) before any definitive statements can be made. The same reservation is also true for adult listeners' perceptions of child talkers. There appears to be minimal evidence that children attempt to mimic the acoustic characteristics of the adult speech that they hear, although we do know from Garnica (1974) that at least the mother is adjusting the acoustic characteristics of her utterances to the child. Why is it then that children's vowel utterances are so clearly delineable at a relatively early age? As discussed by Verbrugge et al. (1976) normalization does not appear to be a satisfactory answer.

References

- DiSimoni, F.G. (1974a): "Evidence for a theory of speech productions based on observations of the speech of children", JASA 56, 1919-1921.
- DiSimoni, F.G. (1974b): "The effect of vowel environment on the duration of consonants in the speech of three-, six-, and nine-year-old children", JASA 55, 360-361.
- DiSimoni, F.G. (1974c): "Influence of consonant environment on the duration of vowels in the speech of three-, six-, and nine-year-old children", JASA 55, 362-363.
- Eguchi, S. and I.J. Hirsh (1969): "Development of Speech Sounds in Children", Acta Oto-Laryngologica 257, 7-51.
- Garnica, O. (1974): Unpublished Ph.D. Diss., Stanford University.
- Gilbert, J.H.V. (1970): "Formant Concentration Positions in the Speech of Children at Two Levels of Linguistic Development", J. Acoust. Soc. Amer. 6,2, 1404-1406.
- Gilbert, J.H.V. (1970): "Vowel Production and Identification by Normal and Language Delayed Children", J. Exper. Child Psych., 9, 12-19.
- Gilbert, J.H.V. (1973): "Acoustical Features of Childrens' Vowel Sounds: Development by Chronological Age Versus Bone Age", Language and Speech 16,3, 218-223.
- Gilbert, J.H.V. (1977): "The Identification of Four Vowels by Children 2½ to 3 Years Chronological Age as an Indicator of Perceptual Processing", In, Segalowitz, S. and F. Gruber (eds.), Language Development and Neurolinguistic Theory, New York: Academic Press, Chapter 19.
- Gilbert, J.H.V. and V.J. Wyman (1975): "Discrimination Learning of Nasalized and Non-Nasalized Vowels by Five-, Six-, and Seven-Year-Old Children", Phonetica 31, 65-80.
- Greenlee, M. (1978): Unpublished Ph.D. Diss., University of California, Berkeley.
- Harrison, G.A., J.S. Weiner, J.M. Tanner and N.A. Barnicot (1964): Human Biology, Oxford: The University Press.
- Kasuya, H., H. Suzuki, and K. Kido (1968): "Changes in Pitch and first three formant frequencies of five Japanese vowels with age and sex of speakers", Research Institute of Electrical Communication, Tokuki University, 344-346.
- Kent, R.D. (1976): "Anatomical and neuromuscular maturation for the speech mechanisms: Evidence from acoustic studies", JSHR 19, 421-447.
- Kent, R.D. (1978): "Imitation of synthesized vowels by pre-school children", J. Acoust. Soc. Amer. 63,4, 1193-1198.
- Lieberman, P. (1978): "On the Development of Vowel Production in Young Children", Paper presented at "Child Phonology, Perception, Production and Deviation", Bethesda, Md. May 28-31.
- Menyuk, P. (1967): "Children's Perception of a Set of Vowels", QPR 84, M.I.T., 254-313.
- Negus, V.E. (1949): "The Comparative Anatomy and Physiology of the Larynx", N.Y.: Hafner.
- Okamura, M. (1966): "Acoustical Studies on the Japanese Vowels in Children", Japanese J. Otol. 69,6, 1198-1214.
- Peterson, G.E. and H.L. Barney (1952): "Control methods used in a study of the vowels", JASA 32, 175-184.
- Verbrugge, R.R., W. Strange, D.P. Shankweiler, and T.R. Edman (1976): "What information enables a listener to map a talker's vowel space", J. Acoust. Soc. Amer. 60,1, 198-212.

## THE CONTROL OF TIMING IN CHILDREN'S SPEECH

Sarah Hawkins, Dental Research Center, University of North Carolina, Chapel Hill, NC 27514, U.S.A.

It is generally acknowledged that temporal and prosodic variables significantly affect speech intelligibility. For example, adult listeners derive considerable information about the syntax and stress pattern of sentences, when segmental cues are either distorted by spectral rotation or absent because the sentence is hummed. The role of prosody in defining syntactic boundaries has been demonstrated with stylised synthetic intonation contours and with prosody pitted against syntax in cross-spliced sentences. The duration alone of both phonemes and larger units can be crucial to speech intelligibility. Additionally, adults appear to be particularly sensitive to the rhythmic onset of stressed syllables, both when listening to speech and when tapping to the rhythm of their own speech. The listener appears to anticipate when stresses will occur and focuses attention at these times. (Documentation of the above points may be found in papers for the other Symposia of this Congress and in Cohen and Nooteboom (1975).) This integrative and predictive role of prosodic cues figures prominently in recent models of speech perception. For example, Pisoni and Sawusch (1975) suggest prosodic cues may form an interface between low-level segmental information and higher levels of syntax and meaning. Martin (1972) has elaborated the notion of the predictive role of rhythm in speech perception, pointing out that efficient perceptual strategies such as attention-cycling between input and output can be facilitated when the signal need not be monitored continuously.

What relevance have these observations about adult speech to children's perception and production of speech? Although the adult listener may be assumed to attend only minimally to much of the acoustic signal, this cannot be assumed for the child. The young child lacks the linguistic and nonlinguistic experience that would allow him/her to "fill in" a large proportion of the message on the basis of knowledge shared with the speaker. Our sparse knowledge of children's perceptual abilities supports this distinction between adults' and children's perception of speech. For

example, although infants of less than 16 weeks can discriminate between stimuli differing in some durational aspects, such as VOT (e.g. Eimas et al., 1971) and syllable duration (Spring and Dale, 1977), children as old as 4-6 years do not necessarily use these durational cues in the same way as adults (Zlatin and Koenigsnecht, 1975; Simon and Fourcin, 1978; Higgs and Hodson, 1978). We know more about children's speech production than their perception; the last 5 or 6 years have provided data on children's timing in phrases, words, syllables (Hawkins and Allen, 1978), segments, and subsegmentals such as VOT (for additional references see below). Many of these studies have demonstrated that while some aspects of children's speech timing resemble those of adult speech from quite an early age, (about 2-4 years), other aspects do not mature until much later (up to about 9-11 years).

The question becomes when and why there are differences between adults and children. Do timing rules appear in children's speech simply as a consequence of increasing neuromuscular coordination, and only gradually come to serve a perceptual function? Or does the child learn the perceptual function of such cues and attempt to produce them in his/her own speech? In the latter case the age when adult timing relationships appear would depend partly on neuromuscular abilities and partly upon the age when their perceptual function is recognised.

This paper discusses ways in which we might distinguish between the above possibilities, using data from children's speech production. The aim is to provide a conceptual framework that will be useful in thinking about children's timing control, as a first step towards formulating a theory of the development of speech timing.

I begin from the position that the child's perception of speech neither is essentially mature before s/he begins to speak (cf. Smith, 1973), nor develops concurrently with production (cf. Waterson, 1970, 1971a,b). Rather, I shall assume that while the young speaker perceives some parts of the speech signal quite maturely, s/he perceives other parts only as unanalysed 'noise'. This view is similar to that of Ingram (1974), except that I assume that the position of the 'noise' may not be fixed in the signal in a segment-by-segment manner. The approach is polysystemic: the child's systems of perception and production both may

be described in terms of quasi-independent subsystems, any or all of which may be in a state of considerable flux at a given time.

I assume also that processes manifested in the child's speech will appear in adult speech and most (but perhaps not all) processes of adult speech will appear in child speech. What distinguishes the two is the domain of influence of each process. Thus in adult speech we may find evidence for both hierarchically integrated "comb" models and sequentially ordered "chain" models of timing (Bernstein, 1967; Kozhevnikov and Chistovich, 1965; Ohala, 1975). Phonemes may be integrated into syllables, for example, consistent with the "comb" model, while the timing of successive higher units may be relatively independent of each other, consistent with a "chain" model. In the young child's speech, such integration implying a "comb" model may not be evident at the syllabic level, but similar integration may occur with less complex units. This reasoning suggests that the child's task in learning to speak fluently is not so much that of learning new routines as of applying similar routines to increasingly more complex domains, thereby integrating the elements of these domains into functional units (c.f. Turvey's (1977) action plans). During this learning, the role played by different processes may change. Auditory feedback in children's speech, for example, appears to have a qualitatively different effect than in adults' speech (e.g. Fry, 1966; MacKay, 1967).

Bearing the above assumptions in mind, let us consider some of the different factors that are likely to affect the speech learning process, together with examples of evidence for their existence within speech timing. Three such factors will be distinguished of which only the second and third are mutually exclusive: (1) processes common to all motor skill development; (2) temporal distinctions that serve as primary perceptual cues; and (3) temporal regularities that do not function as primary perceptual cues.

Processes common to all motor skill development will apply to all aspects of speech development; examples are slower and more variable performance. These phenomena have been demonstrated for speech in many studies and at many levels of analysis from the phrase to the segment (e.g. Eguchi and Hirsh, 1969; DiSimoni, 1974a,b; Tingley and Allen, 1975; Kent and Forner, 1977; Keating and Kubaska, 1978; Smith, 1978; Hawkins, in press). It could be

argued that slower durations occur because the child possesses an articulation-dominant system, whereas the (English-speaking) adult uses a timing-dominant system (cf. Ohala, 1970). Nevertheless, the basis of such articulation-dominance is plausibly immature neuromuscular coordination, requiring more time to achieve adequate articulatory targets. Where a phonological length distinction occurs, the child seems to learn to shorten rather than lengthen articulatory units to produce it. This has been suggested for example for the effect of position-in-utterance on word duration (Keating and Kubaska, 1978), for the development of unstressed syllable production (Allen and Hawkins, in press), and in older children for the effect of clustering on consonant duration (Gilbert and Purves, 1977; Hawkins, in press).

Even in such apparently simple cases as longer duration, a single effect may have more than one underlying cause. For example, measuring /b, d, t/ durations in simple environments, Smith (1978) observed that /t/ was 40% longer in the speech of 2 and 4 year olds than might be expected just on the basis of estimates of the degree of durational increase due to neuromotor immaturity. Smith suggested 3 possible causes for this, any or all of which may have contributed to the observed effect: (i) an effort to increase perceptual differences between /t/ and /d/, (ii) greater complexity of the tongue tip innervation required for /t/ than /d/, and (iii) greater complexity of laryngeal adjustments for voiceless stops over voiced ones.

Temporal distinctions that serve as primary perceptual cues are likely to be detected by the child relatively early as long as they do not signal, for example, syntactic or semantic distinctions beyond the child's comprehension. Hence they should appear in the child's speech in an order reflecting the complexity of the neuromotor coordination required. I shall discuss two examples: phonemically conditioned vowel duration, and voice onset time (VOT) in stops.

Vowel duration functions in English as a primary perceptual cue to the voicing of following consonants, with longer vowels preceding voiced consonants. There is some evidence that this is a distinction that occurs naturally and has been exaggerated in some languages, such as English (Lisker, 1974). Such evidence would suggest that the child might learn to produce the mature

voiced-voiceless ratios relatively early. Naeser (1970) found they were present by 21 months of age and in fact preceded control of the voicing feature that governs the distinction in adult speech.

The development of VOT control in stops nicely illustrates the following points: (i) perception and production do not always develop hand in hand, and (ii) a phonemic distinction that may be legitimately regarded as lying along a single phonological dimension should not necessarily be treated as two extremes of a unitary process in a theory of speech development. The development of the voicing contrast in English has been studied longitudinally (e.g. Kewley-Port and Preston, 1974; Macken and Barton, 1978) and cross-sectionally (e.g. Menyuk and Klatt, 1975; Zlatin and Koenigsnecht, 1976; Gilbert, 1977). It has been consistently found that children make a distinction between short-lag (voiced) and long-lag (voiceless) stops fairly early (around 2 years), but at this stage only the short-lag distribution resembles the adult form. It is not until much later (around 6 years or more) that the long-lag VOT distribution resembles that of the adults. The difference in age of mastery of the two VOT categories is commonly accepted as being due to differences in the neuromuscular coordination required: short-lag stops allow considerable variability in the coordination of laryngeal and oral activity, whereas long-lag stops require rather precise coordination.

Temporal regularities that do not function as primary perceptual cues, especially those that appear to provide no perceptual information, would be expected to be acquired as the child's action plans (articulatory programs) become more sophisticated. An example is the reduction of the duration of consonants in clusters. Although many of the durational differences between clustered and unclustered consonants are perceptible, they may not serve a perceptual function (Klatt, 1976). The age when children produce these durational modifications varies according to the type of cluster, but the last ones are probably not mastered until 9-11 years (Gilbert and Purves, 1977; Hawkins, in press). This is later than many of the temporal phenomena discussed above. Hawkins (in press, in preparation) discusses evidence from these data suggesting that many clusters undergo considerable reorganisation as complex units, rather than there being simply durational

reduction of each segment. Intermediate stages may involve uneven rates of development and durational changes in the opposite direction from that finally required. These observations, together with the late attainment of mature durational relationships, are consistent with the development of increasingly sophisticated motor action plans by the integration of subroutines.

This paper has discussed some of the considerations that should be included in a theory of the development of speech timing within a polysystemic, parallel processing approach. It suggests that adults and children will differ in speech production processes not so much in the nature of those processes as in their relative importance and their domain of influence. The child's 'system' cannot be regarded as static at any time, but rather as reflecting the effects of several continually changing systems that replace each other during development. Changes in one subsystem may affect others, producing either progression or temporary regression. Furthermore, a given phenomenon observed in development may have several causes, whose effects may all work in the same direction or in conflicting directions. Consistent with these points, the development of speech timing may be usefully considered in the contexts of maturing neuromuscular coordination and the perceptual cueing function of timing rules. Neuromuscular immaturity influences patterns of production both across-the-board and in specific contexts. It is reasonable to expect timing rules that are primary perceptual cues to be implemented earlier than those that are not, assuming both that the degree of neuromuscular complexity is constant and that the child perceives the distinction as linguistically relevant.

#### Bibliography

- Allen, G. D. and S. Hawkins (in press): "Trochaic rhythm in children's speech", in Current Issues in the Phonetic Sciences, H. Hollien and P. Hollien (eds.).
- Bernstein, N. A. (1967): The Coordination and Regulation of Movements, Oxford: Pergamon Press.
- Cohen, A. and S. G. Nootboom (1975): Structure and Process in Speech Perception, Berlin: Springer-Verlag.
- DiSimoni, F. G. (1974a): "Effect of vowel environment on the duration of consonants in the speech of 3-, 6-, and 9-year-old children", JASA 55, 360-361.
- DiSimoni, F. G. (1974b): "Influence of consonant environment on duration of vowels in the speech of 3-, 6-, and 9-year-old children", JASA 55, 362-363.



- Eguchi, S. and I. J. Hirsh (1969): "Development of speech sounds in children", AcOtolaryng 257, 1-51.
- Eimas, P., E. Siqueland, P. Jusczyk and J. Vigorito (1971): "Speech perception in infants", Science 171, 303-306.
- Fry, D. B. (1966): "The development of the phonological system in the normal and the deaf child", in The Genesis of Language, F. Smith and G. A. Miller (eds.), 187-206, Cambridge, Mass.: MIT Press.
- Gilbert, J. H. V. (1977): "A voice onset time analysis of apical stop production in three-year-olds", J. Ch. Lang. 4, 103-110.
- Gilbert, J. H. V. and B. A. Purves (1977): "Temporal constraints on consonant clusters in child speech production", J. Ch. Lang. 4, 417-432.
- Hawkins, S. (1973): "Temporal coordination of consonants in the speech of children: Preliminary data", JPh 1, 181-217.
- Hawkins, S. (in press): "Temporal coordination of consonants in the speech of children: Further data", JPh.
- Hawkins, S. (in preparation): "Processes in the development of speech timing control".
- Hawkins, S. and G. D. Allen (1978): "Acoustic-phonetic features of stressed syllables in the speech of 3-year-olds", JASA 63, Suppl. 1, S56.
- Higgs, J. W. and B. W. Hodson (1978): "Phonological perception of word-final obstruent cognates", JPH 6, 25-35.
- Ingram, D. (1974): "Phonological rules in young children", J. Ch. Lang. 1, 49-64.
- Keating, P. and C. Kubaska (1978): "Variation in the duration of words", JASA 63, Suppl. 1, V10.
- Kent, R. D. and L. L. Forner (1977): "A developmental study of speech production: Data on vowel imitation and sentence repetition", JASA 62, Suppl. 1, 0015.
- Kewley-Port, D. and M. Preston (1974): "Early apical stop production: A voice onset time analysis", JPh 2, 195-210.
- Klatt, D. H. (1976): "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence", JASA 59, 1208-1221.
- Kozhevnikov, V. A. and L. A. Chistovich (1965): "Speech: Articulation and perception", Joint Publications Research Service, Washington, D. C.: 30, 543.
- Lisker, L. (1974): "On 'explaining' vowel duration variation", Haskins Labs.: Status Report on Speech Research SR-37/38, 225-232.
- MacKay, D. G. (1967): "Metamorphosis of critical interval: Age-linked changes in the delay in auditory feedback that produces maximal disruption of speech", JASA 43, 811-821.
- Macken, M. A. and D. Barton (1977): "A longitudinal study of the acquisition of the voicing contrast in American-English word-initial stops, as measured by voice-onset time", Stanford University: Papers and Reports on Child Language Development 14, 74-120.
- Martin, J. G. (1972): "Rhythmic (hierarchical) versus serial structure in speech and other behavior", Psych. Rev. 79, 487-509.
- Menyuk, P. and M. Klatt (1975): "Voice onset time in consonant cluster production by children and adults", J. Ch. Lang. 2, 223-231.
- Naeser, M. A. (1970): "The American child's acquisition of differential vowel duration", Madison, Wisconsin: Technical Report 144.
- Ohala, J. J. (1970): "Aspects of the control and production of speech", UCLA Working Papers in Phonetics, 15.
- Ohala, J. J. (1975): "The temporal regulation of speech", in Auditory Analysis and Perception of Speech, G. Fant and M. A. A. Tatham (eds.), 431-453, London: Academic Press.
- Pisoni, D. B. and J. R. Sawusch (1975): "Some stages of processing in speech perception", in Structure and Process in Speech Perception, A. Cohen and S. G. Neebboom (eds.), 16-35, Berlin: Springer-Verlag.
- Simon, C. and A. J. Fourcin (1978): "Cross-language study of speech-pattern learning", JASA 63, 925-935.
- Smith, B. (1978): "Temporal aspects of English speech production: A developmental perspective", JPh. 6, 37-67.
- Smith, N. V. (1973): The Acquisition of Phonology: A Case Study, Cambridge, England: The University Press.
- Spring, D. R. and P. S. Dale (1977): "Discrimination of linguistic stress in early infancy", JSHR 20, 224-232.
- Tingley, B. M. and G. D. Allen (1975): "Development of speech timing control in children", Ch. Dev. 46, 186-194.
- Turvey, M. T. (1977): "Preliminaries to a theory of action with reference to vision", in Perceiving, Acting, and Knowing: Toward an Ecological Theory, R. Shaw and J. Bransford (eds.), Hillsdale, N. J.: Lawrence Erlbaum Associates.
- Waterson, N. (1970): "Some speech forms of an English child: A phonological study", Transactions of the Philological Society, London: 1-24.
- Waterson, N. (1971a): "Child phonology: A prosodic view", JL 7, 179-211.
- Waterson, N. (1971b): "Child phonology: A comparative study", Transactions of the Philological Society, London: 34-50.
- Zlatin, M. A. and R. A. Koenigsknecht (1975): "Development of the voicing contrast: Perception of stop consonants", JSHR 18, 541-553.
- Zlatin, M. A. and R. A. Koenigsknecht (1976): "Development of the voicing contrast: A comparison of voice onset time in stop perception and production", JSHR 19, 93-111.

CROSS-LINGUISTIC EVIDENCE ON THE EXTENT AND LIMIT OF INDIVIDUAL VARIATION IN PHONOLOGICAL DEVELOPMENT

David Ingram, University of British Columbia, Vancouver, B.C., Canada

Let me begin by describing four types of variation that could occur during children's phonological development:

1. intra-child variation: the production of different phonetic forms for the same word, or the inconsistent use of a particular phonological process across words by an individual child;
2. inter-child variation: the production of different phonetic forms or different phonological processes by different children at a comparable stage of development;
3. intra-language variation: the occurrence of different developmental stages or patterns by children learning the same language;
4. inter-language variation: the existence of phonological processes for all children learning a particular sound pattern in one language distinct from those used by all children learning a similar pattern in another language.

We need to determine, first, do all four possibilities actually occur in development? and second, what is the extent and limits of each? Here I will briefly demonstrate that the first three exist, and then comment on what it means to look for inter-language variation.

Intra-child variation, or the use of varying phonetic forms by the same child has been noted for a long time in phonological diaries. Recently, investigators have attempted to document and explain this occurrence. A succinct and plausible description of what may be occurring is expressed in Klein (1977, 159): "It appears then, that the amount of variation in a child's productions may be a function of the type and variety of processes a child has available and applies in modifying words as he/she attempts to say them".

The existence of inter-child variation is also well-documented. Children learning the same words and sounds at comparable periods of development will often show varying ways to produce them, seemingly due to preferences for particular sounds or syllabic shapes, e.g. Priestley (1977). Klein (1977) has dealt with one such pattern at length, that of a preference for reducing syllables versus

one for syllable expansion as in reduplication. Ferguson, in several papers, has referred to such alternatives as individual strategies and suggests that the extent of variation may be quite great. For example, Ferguson and Farwell (1975, 437) in a study on the phonological development of three children during the first 50 words conclude: "each of the three children is exhibiting a unique path of development, with its individual strategies and preferences and an idiosyncratic lexicon".

The occurrence of inter-child variation at any stage implies the existence of intra-language variation, i.e. different developmental stages or paths. Elsewhere, however, (Ingram 1974b, 1978) I have argued that one needs to be cautious about claims of wide-spread variation and alternative stages. Even though variation in the production of specific words and the use of phonological processes may occur, a broader view may indicate that this variation is simply part of a more general pattern. Regarding phonological processes, I suggest that children may follow them in varying degrees (Ingram, 1974b). For variations that result from preference for certain sounds, it may also be that more similar or general patterns are discernible once sound classes are observed (Ingram, 1978).

So far, little research has been done on inter-language variation, with the prevailing position being that most cross-linguistic data will be similar to intra-language findings. In my earlier study of general phonological processes (Ingram, 1974a), I observed that these tended to occur with children acquiring different languages, a point also made in regard to the less obvious process of fronting (Ingram, 1974b). While these claims deal with aspects of development shared by children, similar ones can be found for variations between children. In Ingram et al. (to appear), we studied the acquisition of English word-initial fricatives and affricates across 73 children, determining their individual preferences. Most generally, one could say that when preferences occurred, they were for either labial, alveolar, or palatal productions. This three-way possibility was also observed in Ingram (in press) for three French children, Elie-Paul (Vinson, 1915), Fernande (Roussey, 1899-1900), and Suzanne (Deville, 1890-91) as found in their substitution summarized in table 1.



Table 1

Substitution patterns for French fricatives by three French-learning children

Adult Sound	Substitutions		
	Elie-Paul (1;11)	Fernande (2;4)	Suzanne (1;11)
/f/	-	s	f
/v/	-	s	v
/s/	ʃ	s	s(t)
/z/	ʃ	-	z
/ʃ/	ʃ	s	s
/ʒ/	ʃ	s	ʒ

Is there, then, inter-language variation, i.e. the existence of distinct phonological processes for learners of one language that do not occur for learners of another language? To begin, there are two simplistic extremes that reduce the issue to a trivial one. On the one side, we can say that children are all genetically prepared to learn language, and consequently all have the same processing mechanisms available. In this case, no inter-language variation is possible, for it may simply be that the proper circumstances for a particular process do not apply. For example, a child will not show simplification of consonant clusters in learning a language that does not have them. The other situation is to say that all languages are phonologically different, so that of course children will show differences across languages because there are distinct phonetic inventories and phonological patterns to be learned.

There is, however, a middle ground between these two where variation may be viewed in a non-trivial fashion. This is where there are cases when two languages appear to present children with similar phonological patterns, but children do not deal with them in the same way. A closer analysis should provide insight into the multiple conditions that occur in a particular language which lead to different learning patterns across languages.

Let me provide two examples of such patterns. In English, there is a common process of velar assimilation where an alveolar consonant will assimilate to a following stop if it is velar, e.g.

Jennika 1;7 duck [gʌk]; 2;2 tickle [gigu]. This occurs in both CVCs where the sounds are within a syllable, and in CVCV words where the assimilating sounds cross a syllable boundary. I have examined extensive data from the three French children mentioned above and have not found instances of this process. Possible explanations are that potential instances are rare, and that the different timing pattern of French inhibits its occurrence. Nonetheless, one can locate places where it could occur, given our current understanding of the process.

A second example concerns a process that all three French children show quite widely, but which is not found in English learning children with any frequency. This is the process of consonant denasalization where a nasal consonant will denasalize in harmony with a nonnasal obstruent, e.g. Fernande mange 'eat' [baʃ]; menton 'chin' [ba:to:]; marcher 'walk' [base]. Table 2 presents data indicating how dominant the pattern is for the three French children under discussion.

Table 2

Proportion of occurrence of denasalization for three French children at varying ages

age	Suzanne	age	Fernande	age	Elie-Paul
1;0-1;7	.00 (0/1)	1;4-1;9	1.00 (5/5)	2;1	.50 (2/4)
1;8	.67 (4/6)	1;10-2;0	.83 (5/6)	2;2-2;5	1.00 (7/7)
1;9	.67 (6/9)	2;1-2;3	1.00 (8/8)	2;6-3;0	.25 (2/8)
1;10	.50 (6/12)	2;4-2;7	.25 (1/5)		
1;11	.43 (6/14)	2;8-2;10	.00 (0/7)		
2;0	.19 (3/16)				

Like the previous example, one can cite some factors that might contribute to its nonoccurrence in English, but potential cases do arise.

Data like these suggest that inter-language variation occurs, and that we need to seek more instances of it. Once more are found, they should show that phonological acquisition is the complex interaction of several conditions in the adult language, and that phonological processes will need to be described in much more detail

than exists to date. Also, they indicate that the study of acquisition in only one language may yield a restrictive, and possibly misleading, view of the language learning process.

#### References

- Deville, G. (1890-91): "Notes sur le développement du langage", Revue de Linguistique et de Philologie Comparée 23, 330-343; 24, 10-42, 128-143, 242-257, 300-320.
- Ferguson, C. and C. Farwell (1975): "Words and sounds in early acquisition", Lg 51, 419-439.
- Ingram, D. (1974a): "Phonological rules in young children", Journal of Child Language 1, 49-64.
- Ingram, D. (1974b): "Fronting in child phonology", Journal of Child Language 1, 233-241.
- Ingram, D. (1978): "The acquisition of fricatives and affricates in normal and linguistically deviant children", in The Acquisition and Breakdown of Language, A. Caramazza and E. Zuriff (eds.), 63-85, Baltimore: Johns Hopkins University Press.
- Ingram, D. (in press): "Phonological patterns in the speech of young children", in Studies in Language Acquisition, P. Fletcher and M. Garman (eds.), Cambridge: Cambridge University Press.
- Ingram, D., L. Christensen, S. Veach and B. Webster (to appear): "The acquisition of word-initial fricatives and affricates in English by children between two and six", in Child Phonology: Data and Theory, J. Kavanaugh, G. Yeni-Konshian, and C. Ferguson (eds.).
- Klein, H. (1977): The Relationship between Perceptual Strategies and Productive Strategies in Learning the Phonology of Early Lexical Items, Doctoral dissertation, Columbia University.
- Priestley, T.M.S. (1977): "One idiosyncratic strategy in the acquisition of phonology", Journal of Child Language 4, 45-66.
- Roussey, C. (1899-1900): "Notes sur l'apprentissage de la parole chez un enfant", La Parole 1, 870-880; 2, 23-40.
- Vinson, J. (1915): "Observations sur le développement du langage chez l'enfant", Revue de Linguistique 49, 1-39.

THE ACQUISITION OF CHINESE PHONOLOGY IN RELATION TO JAKOBSON'S  
LAWS OF IRREVERSIBLE SOLIDARITY

Heng-hsiung Jeng, National Taiwan University, Taipei,  
Republic of China

I. Introduction

This paper attempts to find out whether the laws of irreversible solidarity as proposed by Jakobson (1968; 1971) also apply to the acquisition of Chinese phonology by two Chinese children.

Chinese here refers to the Mandarin Chinese as spoken in Taiwan, Republic of China, today. This variety of Mandarin Chinese is different from the standard Mandarin in that the former generally does not have the retroflex affricates /tʂ/, /tʂ<sup>h</sup>/, and the retroflex fricative /ʂ/ that can be found in the latter. As far as tones and other segmental phonemes are concerned, they are essentially the same.

The subjects are my first son Jeng Wei, born on October 15, 1969, and my second son Jeng Hung, born on June 5, 1975. The data selected for this study are those of my first son recorded between the age of 2 months when babbling started and the age of 20 months when I left for the U.S. and stopped recording, and those of my second son recorded between the age of 15 months when he began to utter the first words and the age of 31 months when he had more or less mastered Chinese phonology. All these data were recorded by the author, mainly in phonetic transcription and occasionally with a tape recorder. My first son's data were mainly used for the discussion of the acquisition of tones, and the second son's data for the discussion of the acquisition of segmental phonemes.

My wife's native Chinese dialect is Hakka, and my native Chinese dialect is the South Min dialect spoken in Taipei. Most of the time we converse in the variety of Mandarin Chinese spoken in Taiwan as characterized above. Only when my wife's relatives or mine come to visit us is Hakka or South Min heard more often. So my sons generally live in the native-speaking environment of Mandarin Chinese, with only occasional exposure to Hakka and South Min.

II. Acquisition of Chinese Phonology

A. Tones

It has been observed by Jakobson (1968, 21-22) and Lenneberg (1964, 119) that babbling is not directly related to the acquisi-

tion of speech sounds. However, they did not touch upon the acquisition of tones in a tone language.

Chao (1951) noted that most Chinese children acquire tones quite early, except some tone sandhi phenomena. Li and Thompson (1976) and Li (1978) also observed that the acquisition of tones by a Chinese child precedes the acquisition of segmental phonemes. Weir (1966, 156) even pointed out that a Chinese baby at about six months, that is, during its babbling stage, had already much tonal variation over individual vowels, while the Russian and American babies at about the same age seldom showed such variation.

The written records of my first son's babbling show that he had tonal variation over individual vowels or syllables at a very early age: [ə/ē] (2 months); [e/ē] (3 months); [ɿ/ɿ̃] (3 months); [kuẽkuẽ] (3 months). And at 4 months, in response to my utterance [ãã], he produced a similar tonal variation over the same vowels. The above examples show that at the very early stage of babbling, he not only could produce vowels and syllables with different tones, but also link such production to the perception of tones.

This evidence supports Chao's (1951), Li and Thompson's (1976) and Li's (1978) observations that tones are acquired quite early by Chinese children. With such ability to perceive and produce tones transferred from the babbling stage, a Chinese child usually sets out to acquire his first Chinese words with practically correct tones. My first son, at 11 months, uttered his first word [papa] correctly with a falling tone followed by a neutral tone. At one year, he produced the word [pa<sup>w</sup>] 'to mix milk powder with water' correctly with the falling tone even though the aspirated voiceless bilabial stop /p<sup>h</sup>/ was incorrectly pronounced as its unaspirated counterpart. At 15 months, he could recognize the difference between [ɕiẽiẽ] 'shoes' and [ɕiẽiẽ] 'thanks', [xua] 'flower' and [xua] 'painting' because of their different tone patterns, even though he could not produce them yet. My second son at 16 months, about one month after the utterance of his first word, produced a minimal pair with tones as the distinctive elements: [pa<sup>w</sup>pa<sup>w</sup>] 'bread; food' and [pa<sup>w</sup>pa<sup>w</sup>] 'hold in arms'.

Mandarin Chinese has four tones and one neutral tone, which occurs only in an unstressed syllable. Besides the above mentioned high level tone [˥] (55), namely the first tone, in such a word

as [xua] 'flower', falling tone [˨˨˨] (51), namely the fourth tone, in such a word as [xua] 'painting', and neutral tone [˧˧˧] in the second syllable of the word [papa] 'father', there are the rising tone [˨˨˨] (35), namely the second tone, and the falling-rising tone [˨˨˨] (214), namely the third tone, which is realized as [˨˨˨] (35) when it occurs immediately before another third tone and normally realized as [˨˨˨] (21) elsewhere. Both my first and second sons acquired the second and third tones more or less simultaneously and without much difficulty: my first son had been able to produce the second-tone words [mai] 'buy', [niu] 'cow', and the third-tone words [tɕi] 'self' in [tɕi tɕi lai] 'by oneself', [tɕi] 'rise' in [tɕi lai] 'get up' by the age of 19.5 months; my second son uttered the second-tone words [tɕien] 'money' at 17.5 months, [nai] 'come' at 18.5 months, and the third-tone words [ta kai] 'open' at 16.5 months, [pa pe] 'urinate' at 17 months. However, they occasionally mispronounced some second-tone words as third-tone words and vice versa, and this supports the view of Li and Thompson (1976, 189) concerning such occasional confusion.

As for tone sandhi phenomena, the data of my second son, contrary to Chao's (1951) observations, show that he generally had no problem with them. And this also supports the view of Li and Thompson (1976, 189) that "tone sandhi rules are learned, with infrequent errors". For example, the third-tone word /uo/ 'I' before another third-tone word was correctly changed to the second tone in the expression [uo ie iau tsu tsy] 'I also want to go out' uttered at 21.5 months, while before a neutral-tone word, it was correctly realized as [uo] in the expression [uo tɕ] 'mine', uttered at 21.5 months. And the fourth-tone word /pu/ 'not' before another fourth-tone word was correctly changed to the second tone in the expression [pu tsai] 'absent' uttered at 24 months, but before a third-tone word, it remained unchanged in the expression [pu ɕi xuan] 'don't like' uttered at 22.5 months.

Therefore, Chinese tones, unlike segmental phonemes which have to evolve slowly step by step, are perhaps acquired by a Chinese child during babbling before the utterance of the first word and assigned immediately to the first words acquired.

But why are tones acquired before segmental phonemes? Perhaps the answer may be found in the lateralization of the human brain. Fromkin and Rodman (1974, 312) state that after lateralization,

the right brain is specialized in pattern-matching and the left brain in analytical thinking. Probably that is why such discrete linguistic elements as segmental phonemes can be acquired by the left brain after lateralization, which takes place around the age of one, and before lateralization, when both sides of the brain are still symmetrical, only suprasegmental patterns such as tones can be acquired.

#### B. Consonants and Vowels

My second son's acquisition of Chinese segmental phonemes may be divided into two stages: i. minimal phonological system; ii. fully developed phonological system, which is almost identical with an adult Mandarin speaker's phonology.

The minimal phonological system consists of four stops, /p/, which has two allophones [m] and [b] as free variants, /t/, /k/, and /ts/, which has an allophone [tʂ] occurring before /i/, and four vowels, /a/, /a<sup>w</sup>/, /i/, and /e/. All these segmental phonemes were acquired within 44 days, between September 25 and November 8, 1976. And the words uttered within this period are as follows:

(9/25)<sup>1</sup> [pa<sub>q</sub>pa<sub>q</sub>] 'people'; (9/26) [ka<sup>w</sup>ka<sup>w</sup>] 'older brother', [ie tɕi]~[te tɕi] 'eyes'; (10/4) [a tɕitɕi] 'dirty'; (10/12) [pa<sup>w</sup>pa<sup>w</sup>] 'bread; food'; (10/18) [pa<sup>w</sup>pa<sup>w</sup>] 'hold in arms'; (10/19) [tsatsa] 'dirty'; (10/26) [ta kai] 'open', [tata] 'candy', [te ta] 'fall down', [mapa]~[baba]~[papa] 'people'; (10/28) [piapia] 'don't want'; (10/29) [pa?pa] 'car'; (11/2) [pa] 'flower'; (11/7) [tia] 'drop'; (11/8) [pa pe] 'urinate'.

Beyond this stage of minimal phonological system, nasals, aspirated stops, fricatives except /f/, and the retroflex liquid /r/ emerged almost simultaneously although they became stable at different times. The lateral liquid /l/ appeared later than all these sounds, and /f/ was the last sound to appear. The following table shows when these consonants first emerged and when they became stable. The first number under each consonant indicates the age (in months) when it emerged, and the second number the age when

- 
- (1) Hereafter, the Arabic numeral before a slash indicates the month and the Arabic numeral after it indicates the day of the month.
  - (2) This voiced bilabial fricative [β], which evolved into /x/ later on, does not fit into the minimal phonological system proposed here.

it became stable.

Table 1

Emergence and stabilization of further consonants in the fully developed phonological system

	p <sup>h</sup>	t <sup>h</sup>	k <sup>h</sup>	ts <sup>h</sup>	m	n	ŋ	r	f	s	x	l
Emer.	19	21.5	17	20	17	17	19	17	29	18	18	20.5
Stabi.	22.5	22.5	22	22	17	23	22.5	17	29	18	18	20.5

The vowels that emerged in the fully developed phonological system are /u/, /y/, /ɿ/, and /o/, which has the allophone [ɤ] when occurring after a nonlabial sound or occurring as a word-initial vowel. Once these vowels were acquired, they were very stable afterwards, except /y/, which at one time lapsed into [i<sup>w</sup>] for the word "fish", whose proper pronunciation is [y]. The ages when these vowels appeared are given in the following table.

Table 2

Emergence of further vowels in the fully developed phonological system

	u	ɿ	y	o	ɤ
Emer.	17.5	20.5	19.5	18	20

The division of Jeng Hung's phonological development into the minimal phonological system and the fully developed phonological system may appear to be rather arbitrary. However, because of the simple distinctive features involved in the minimal phonological system and the complex distinctive features involved in the fully developed phonological system, the division is not without justification: in the minimal system, each of the consonants has only two distinctive features, that is, [+stop] and point of articulation, and each of the vowels is distinguished from the other vowels by two features, [±high] and [±low], except /a<sup>w</sup>/, which has an additional feature of [+labialized]; in the fully developed system, after the age of 17 months, more consonants are distinguished by

more complex manner features such as [+aspirated], [+nasal], [±fricative], [±liquid], and [±retroflex] even though their points of articulation remain more or less the same as those of the stops in the minimal system, and vowels are further distinguished by [±back] and [±round].

### III. Jakobson's Laws of Irreversible Solidarity

Jakobson (1968; 1971) set forth the laws of irreversible solidarity to account for the chronology of the acquisition of speech sounds by children, sound changes, and loss of speech sounds by aphasics. Now the acquisition of Chinese phonology by my sons Jeng Wei and Jeng Hung will be discussed in the light of his laws.

1) Jakobson did not touch upon the acquisition of tones in tone languages. According to Li and Thompson (1976) and Li (1978), the acquisition of tones precedes the acquisition of segmental phonemes. The discussion in II.A further points out that babbling has an important bearing on the acquisition of tones.

2) In the minimal phonological system of Jeng Hung, the vowels /i/, /e/, and /a/ form a vertical split as Jakobson predicted, but the labialized vowel /a<sup>w</sup>/, which developed into the diphthong /au/ at 17.5 months, does not fit neatly into the pattern, and the consonants /p/, /t/, /k/, and /ts/ deviate from his laws of first and second consonantal split.

3) The early appearance of /k/ in Jeng Hung's minimal phonological system and /k<sup>h</sup>/ and /x/ in his fully developed phonological system forms a counterexample to Jakobson's law that back consonants presuppose front consonants.

4) Jakobson's law that back rounded vowels presuppose their corresponding front unrounded vowels is supported by Jeng Hung's early acquisition of /i/ and /e/ and late acquisition of /u/ and /o/. So is his law that /y/ presupposes /i/ and /u/.

5) The almost simultaneous appearance of the aspirated stops, nasals, fricatives except /f/, and the retroflex liquid /r/ cannot be accounted for by Jakobson's laws. One tentative explanation proposed here is that these aspirated stops, nasals, and fricatives except /f/, being identical with their corresponding unaspirated stops in the minimal phonological system with respect to points of articulation, are developed simultaneously on the basis of adding to the existent unaspirated stops one more distinctive feature from the different manners of articulation such as [+aspirated],

[+nasal], and [+fricative]. The late acquisition of /f/, in the light of this explanation, may be due to the fact that its point of articulation is different from any of the unaspirated stops in the minimal phonological system, hence the substitution of /f/ by the voiceless bilabial fricative [ɸ] in the words [i<sub>1</sub> ɸu<sub>1</sub>] 'clothes' and [ɸei<sub>1</sub> tɕi<sub>1</sub>] 'aeroplane'. As for the simultaneous acquisition of /r/ with aspirated stops, nasals, and fricatives except /f/, one possible explanation is that the additional distinctive feature of [+retroflex] is combined with the negative values of these manners of articulation as a clear-cut opposition.

6) Jakobson (1968) pointed out that the second liquid is one of the last sounds acquired by the child. The late acquisition of /l/ by Jeng Hung at 20.5 months, only before /f/, the last sound acquired, supports his view.

### References

- Chao, Y.R. (1951): "The Cantian idiolect", Semitic and Oriental studies presented to William Popper, University of Calif. Publications in Semitic Philology II, 27-44.
- Fromkin, V.A. and R. Rodman (1974): An introduction to language, New York: Holt, Rinehart and Winston.
- Jakobson, R. (1968): Child language, aphasia and phonological universals, The Hague: Mouton.
- Jakobson, R. (1971): Studies on child language and aphasia, The Hague: Mouton.
- Lenneberg, E.H. (1964): "Speech as a motor skill with special reference to non-aphasic disorders", in The acquisition of language, U. Bellugi and R. Brown (eds.), 115-127.
- Li, C.N. and S.A. Thompson (1976): "The acquisition of tone in Mandarin-speaking children", Journal of Child Language 4, 185-199.
- Li, P.J.K. (1978): "Child language acquisition of Mandarin phonology", in Studies and essays in commemoration of the golden jubilee of the Academia sinica, 615-632.
- Weir, R.H. (1966): "Some questions on the child's learning of phonology", in The genesis of language, F. Smith and G.A. Miller (eds.), 153-172, Cambridge: The MIT Press.

PREDISPOSITIONS FOR THE PERCEPTION OF SPEECH BY HUMAN INFANTS  
Patricia K. Kuhl, Department of Speech and Hearing Sciences,  
 Child Development and Mental Retardation Center, University of  
 Washington, Seattle, WA. 98195.

The development of speech production and perception in the human infant shares certain themes with the acquisition of communicative repertoires in animal species. Among those themes is the notion that infants of a species demonstrate predispositions for the perception of communicatively relevant acoustic signals. While the animal literature provides examples in which innate predispositions are in evidence, a growing body of literature on the complex role of "normal" experience, and the effects of selective auditory exposure, in maintaining, facilitating, and inducing such behavior is accruing, leading to the hypothesis that infants are predisposed toward fairly simple acoustic features and develop the perception of "configurational" models only with experience. Two approaches to examining the role of experience in the perception of speech by human infants are discussed.

#### Converging Themes in Developmental Neurobiology

At the end of the first decade of research on the perception of speech by young infants, the list of published experiments is long and the speech features that have been examined is extensive (see Kuhl, In Press, for review). The common theme running through this work is the examination of potential auditory perceptual predispositions that human infants bring to the task of learning language - predispositions that would direct the infant toward the acoustic features that are particularly relevant to the perception of speech, such as those acoustic features which signal the segmental and nonsegmental elements of the language.

The notion that members of a species may be perceptually predisposed to attend to, resolve more precisely, respond to, or to otherwise treat differently, visual and auditory signals that are relevant to their survival is an old theme in the literature on communicative behavior in animals and humans (Lorenz, 1965). Many attribute stimulus prepotencies to species-specific neural mechanisms that have evolved specially for that purpose and perceptual predispositions that are innately determined. The evidence for such mechanisms is both behavioral and physiological

and largely stems from work on animal communication (see Schneich, 1977, for a review of neurophysiological data and Gottlieb, 1976a, for a review of behavioral data).

The discovery in behavioral and physiological experiments that communicatively relevant stimuli enjoy special status for the adult perceiver naturally raises questions about the development of these behaviors in infants of the species. While the early theorists (Lorenz, 1965; Tinbergen, 1951) stressed the "instinctiveness" of certain behaviors and underplayed the role of experience, "learning" in the classical sense, or maturation in the development of complex behavior, more recent theorists (Gottlieb, 1976a) have stressed the complex role that experience plays and the variety of different ways experience affects the organism (Gottlieb, 1976b).

Recent physiological evidence suggests that sensory input during early development has an effect on central neural mechanisms, particularly in the visual system; the responsiveness of units in the visual cortex of adults is biased by distorting or denying early "normal" visual experience, or by selective visual exposure. This physiological "plasticity" in the visual system can be species-specific and evidence for "critical periods" exists (see Daniels and Pettigrew, 1976, for review).

The effects of selective auditory exposure are less well known. Silverman and Clopton (1977) and Clopton and Silverman (1977) noted substantial losses in binaural interaction at the inferior colliculus in rat after early monaural deprivation. Clopton and Silverman (1978) demonstrated changes in the latency and duration of neural responses to clicks at the level of the inferior colliculus in rat after early auditory deprivation. Clopton and Winfield (1976) further demonstrated using the rat that exposure during the first four months of life to patterned sound (upward tone sweeps, downward tone sweeps, or noise bursts) increases the response of units in the inferior colliculus to that pattern relative to a similar but inexperienced pattern. No effects of selective exposure were found in an adult population of rats.

Perhaps the best examples from the animal literature on the interactions between innate predispositions and experience are



to be found in the growing literature on song learning in the Passerine bird (Marler, 1973). Certain songbirds must hear their songs in order to learn them but there are interesting constraints on learning; the exposure must be to the conspecific song and it must occur during a "critical period." Marler hypothesizes that song vocalization is developed by reference to an "auditory template," a mechanism that is specific enough to detect some of the critical features of the conspecific song and thus direct the bird's attention in its direction, but one which requires exposure to the song to "fill in" the details of its acoustic structure. As Marler (1973) describes, their learning is not left purely to chance, it ". . . takes place within a set of constraints which seem designed to ensure that the learning bird's attention shall be focused on a set of sounds that is biologically relevant. . ." (p.80). To make the songbird parallel even more striking, Nottebohm *et al.* (1976) have demonstrated functional hemispheric asymmetry for the production of song in these birds. Using ablation techniques, they have demonstrated that the left hemisphere controls song production in the Canary, but if ablation of the left motor area occurs before the bird has passed the critical period for vocal learning, the bird's song develops normally using the subordinate right motor area.

Marler interprets these data as indicating that the innate direction that the infant comes into the world with is simply that - a direction or guideline pointing the infant in the appropriate direction, rather than a complete "schema" of the song. He believes that the predispositions are toward rather simple stimulus features and only with continued exposure to the configuration that is being detected does the infant develop a "schema" of the complex stimulus array.

#### Predispositions for the Perception of Speech by Human Infants

There are two ways in which the role of experience is currently being examined for the perception of speech by human infants. One approach is to chart the course and examine the nature of perceptual changes that occur as a result of exposure to a particular language. Another approach is to examine the infant's recognition of abstract auditory-phonetic categories rather than simple stimulus features, expecting that the former

may reveal developmental trends.

How does linguistic exposure modify the way in which infants perceive speech sounds? While not well understood, the perceptual effects of exposure to one's native language have been documented in adult listeners (Miyawaki *et al.*, 1975; Abramson and Lisker, 1970). Taken together with the existing data on the perception of speech by infants, these data have led to the hypothesis that infants discriminate all of the simple phonetic contrasts at birth regardless of their linguistic environments, but that due to the lack of exposure to certain phonetic units during development the infant somehow loses the ability to distinguish them from contrasting phonetic units.

Attempting to chart developmental changes in an infant's perception that can be attributed to linguistic exposure has received some attention, but we are still without a simple answer to the question. The evidence is fairly convincing that infants being reared in non-English-speaking environments are capable of discriminating at least one phonetic contrast (voiceless-unaspirated /pa/ from voiceless-aspirated /p<sup>h</sup>a/) that is phonemic in English but not in the infant's native language. Streeter (1976) using the sucking-habituation technique, demonstrated that two-month-old African Kikuyu infants discriminated the English contrast in addition to discriminating a voicing contrast that is phonemic in the Kikuyu language but not in English (prevoiced /ba/ from voiceless-unaspirated /pa/). Lasky, Syrdal-Lasky and Klein (1975) demonstrated similar results for Spanish infants of the same age using a heart-rate technique.

On the other hand, the case for discrimination of the prevoiced /ba/ from the voiceless-unaspirated /pa/ by American infants is not quite as clear. Recent studies (Eilers, Wilson and Moore, 1977; Eimas, 1974) have failed to provide evidence that American infants discriminate pairs of stimuli that are as close on the continuum as those discriminated by the Spanish and Kikuyu infants. However, there are a number of problems with these cross-language comparisons. First, the stimuli are synthesized to manipulate an acoustic cue that is acoustically fragile and is likely to be subject to variation due to the differences in acoustic calibration across laboratories. A more recent set of



studies claims to be immune to this criticism. Using the head-turn technique, Eilers, Gavin and Wilson (In Press) tested six-month-old American and Spanish infants in the same laboratory, but in two different studies, and demonstrated that while both groups discriminated the English contrast, only the Spanish infants discriminated the Spanish contrast.

Do infants recognize the configurational properties of phonetic categories? Only recently have researchers attempted to find out whether infants are capable of recognizing the similarity among sounds that have the same phonetic label when the sounds occur in different phonetic contexts, when they occur in different positions in a syllable, or when they are spoken by different talkers.

A conditioned head-turn response for visual reinforcement has been successfully used with six-month-old infants to test the recognition of phonetic categories (Kuhl, 1978). In these tasks, infants are trained to make a head-turn response when one speech token is changed to another speech token (like from /a/ to /i/). During training, vowels produced by a male talker (computer-synthesized) are used; subsequently, infants are tested with computer-synthesized vowels produced by female and child talkers. The ease with which the infant generalizes to new exemplars from the category indicates the degree to which the infant perceives the similarity among the tokens from a given category.

Results to date in these category-formation tasks strongly suggest that vowel categories are readily perceived by the infant listeners. Tasks requiring the infant to recognize a change from the vowel category /a/ to the vowel category /i/ and tasks requiring the infant to recognize a change from the vowel category /a/ to the vowel category /ɔ/ result in near perfect transfer of learning to the new tokens from the categories (Kuhl, 1978). We have also completed studies on the categorization of fricative consonants, such as /f/ vs. /θ/, and /s/ vs. /ʃ/ (Holmberg, Morgan and Kuhl, 1977). In general, our results suggest that the /a-i/ contrast is the easiest in this category-formation task, that the /f-θ/ contrast is the most difficult one, and that the /a-ɔ/ and the /s-ʃ/ contrasts are of intermediate difficulty.

These category-formation experiments (discussed in detail in Kuhl, 1978) have two advantages. First, one can test the infant's recognition of abstract configurational properties of speech-sound categories, and second, one can test how readily or efficiently the infant forms categories based on dimensions that are not phonetically relevant, at least in English, such as pitch contour or stress. These techniques may demonstrate that all infants recognize categories based on certain "focal" auditory dimensions, but that their tendencies to attend to particular acoustic dimensions is modified by exposure to a particular language.

Systematic experiments examining the perception of abstract perceptual categories, rather than simple discriminations, in at least two different populations in which the target acoustic features are chosen such that they are phonemically relevant to one population and not to the other are necessary before the contributions of innate predispositions and experience will be understood in the development of speech perception.

#### References

- Abramson, A. and Lisker, L. (1970): "Discrimination along the voicing continuum: Cross-language tests," in Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967, Academic, 569-573.
- Clopton, B.M. and Silverman, M.S. (1977): "Plasticity of binaural interaction. II. Critical period and changes in midline response," J. Neurophysiol. 40, 1275-1280.
- Clopton, B.M. and Silverman, M.S. (1978): "Changes in latency and duration of neural responding following developmental auditory deprivation," Exp. Brain Res. 32, 39-47.
- Clopton, B.M. and Winfield, J.A. (1976): "Effect of early exposure to patterned sound on unit activity in rat inferior colliculus," J. Neurophysiol. 39, 1081-1089.
- Daniels, J.D. and Pettigrew, J.D. (1976): "Development of neuronal responses in the visual system of cats," in Neural and Behavioral Specificity, Vol. III, G. Gottlieb (Ed.), 196-232, New York: Academic Press.
- Eilers, R.E., Gavin, W.J. and Wilson, W.R. (In Press): "Linguistic experience and phonemic perception in infancy," Child Develop.
- Eilers, R.E., Wilson, W.R. and Moore, J.M. (1977): "Speech discrimination in the language-innocent and the language-wise: A study in the perception of voice-onset time," J. Acoust. Soc. Am. Suppl. 1, 61, S38(A).

- Eimas, P.D. (1974): "Linguistic processing of speech by young infants," in Language Perspective-Acquisition, Retardation and Intervention, R.L. Schiefelbusch and L.L. Lloyd (Eds.), 55-74, Baltimore: University Park Press.
- Gottlieb, G. (1976a): "Early development of species-specific auditory perception in birds," in Neural and Behavioral Specificity, Vol. III, G. Gottlieb (Ed.), 237-281, New York: Academic Press.
- Gottlieb, G. (1976b): "The roles of experience in the development of behavior and the nervous system," in Neural and Behavioral Specificity, Vol. III, G. Gottlieb (Ed.), 25-54, New York: Academic Press.
- Holmberg, T.L., Morgan, K.A. and Kuhl, P.K. (1977): "Speech perception in early infancy: Discrimination of fricative consonants," J. Acoust. Soc. Am., Suppl. 1, 62, S99(A).
- Kuhl, P.K. (1978): "Perceptual constancy for speech-sound categories in early infancy," Chapter presented at the NIH Conference on Child Phonology, May, 1978.
- Kuhl, P.K. (In Press): "The perception of speech in early infancy," in Speech and Language: Research and Theory, N.J. Lass (Ed.), New York: Academic Press.
- Lasky, R.E., Syrdal-Lasky, A. and Klein, R.E. (1975): "VOT discrimination by four- to six-and-a-half-month-old infants from Spanish environments," J. Exp. Child Psych. 20, 215-225.
- Lorenz, K. (1965): Evolution and Modification of Behavior, University of Chicago Press, Chicago.
- Marler, P. (1973): The Clarence M. Hicks Memorial Lectures for 1970, University of Toronto Press, Toronto, 69-85.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins J., and Fujimura, O. (1975): "An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English," Percept. Psychophys. 18, 331-340.
- Nottebohm, F., Stokes, T.M. and Leonard C.M. (1976): "Central control of song in the Canary," J. of Comp. Neurol. 165, 457-486.
- Schneich, H. (1977): "Central processing of complex sounds and feature analysis," in Recognition of Complex Acoustic Signals, T.H. Bullock (Ed.), 161-182, Berlin: Abakon Verlagsgesellschaft.
- Silverman, M.S. and Clopton, B. (1977): "Plasticity of binaural interaction. I. Effect of early auditory deprivation," J. Neurophysiol. 40, 1266-1274.
- Streeter, L.A. (1976): "Language perception of 2-month-old infants shows effects of both innate mechanisms and experience," Nature 259, 39-41.
- Tinbergen, N. (1951): The Study of Instinct, Clarendon Press, Oxford.

TRANSITION AND VARIATION IN CHILD PHONOLOGY: MODELING A  
DEVELOPING SYSTEM

Lise Menn, Aphasia Research Center, Boston University  
School of Medicine, Boston, Massachusetts, U.S.A.

Child phonology is different from general phonology in several important areas. When we try to characterize those differences we find in many cases that a set of phenomena which play a central role in the one field play a marginal role in the other. It is quite reasonable a priori that this should be the case when we consider the topic of variation in child phonology versus the topic of variation in adult phonology: the very notion of acquisition implies long-term change in performance, whereas we assume that in the adult, the phonology is sufficiently stable for any change to be relegated to the limbo of marginal phenomena.

In this paper, I will briefly review certain types of variation which are prominent in child phonology, and consider how one might incorporate these types of variation in a theoretical model. We will not take up those types of variation that are prominent in both child phonology and adult phonology, such as registral, sociolinguistic, allomorphic, and allophonic variation, although a complete model must deal with those as well; we will keep to the more restricted topic of those types of variation that seem to be intimately associated with the process of the acquisition of phonology.

These will include, as mentioned, long-term changes in rules and pronunciations. These are orderly, one-way transitions in language behavior: the child learns to hit a particular phonetic target, or learns to render a particular sequence of sounds in accord with the adult model word instead of producing it in some scrambled order.

Acquisition studies show that there are also several types of short-term variation among renditions of a given word. Two of these can be considered as being the microstructure of long-term variation: transitional variation and local scatter in the production of a particular phone in a phonologically defined context.

Transitional variation refers to the vacillation between

well-defined pronunciations of a word that frequently occurs during the period when an old rule is being superseded by a new rule. Such bimodal variation in renditions of a word is usually taken as evidence that two rules are in conflict. Sometimes the changeover from old to new rules has an intermediate period showing transition variation, and sometimes no such period is observed.

Local scatter is a unimodal variability in the production of a particular phone. This simply looks like the result of poor articulatory control compared to the adult norm: the child's shots at a target more often fall wide of the mark. (There must also be a second-order long-term variation associated with local scatter, since we expect to see a reduction in local scatter as the child matures.)

Presently I can enumerate five other kinds of short-term variation. One of these is called backgrounding (Ferguson & Farwell 1975). As they say, one portion of a word may be "deleted or drastically reduced while the child is 'working on' another part of the word." They cite from their data one child's production of 'milk' as [bʌʔ] and [ʌk̄] in the same session. I think we now have enough evidence from selective avoidance (Ferguson & Farwell 1975) to assert that children can and sometimes do monitor the quality of their own output; therefore, the most reasonable explanation of backgrounding as Ferguson & Farwell describe it is to assume that it takes place under conditions of high self-monitoring of the phonetics of the output or the input.

A second type of variation which also seems to involve self-monitoring is the well-documented imitation effect: a word may be pronounced very differently when it is an imitation than when it is produced without the adult model ringing in the child's ears. Frequent anecdotes report one sub-type of model-induced variation: a child will be reported to have said a word 'perfectly' or nearly so on the very first attempt, and then to have reduced it drastically in later renditions. One would expect to find parallels to backgrounding and imitation-effect variability in adult speech when one is attending to the sound of the word as well as its meaning, while speaking.

(It is also well-known that children can spectacularly fail to be aware of the sound of their output, and imitation may fail

to induce any variation at all; he or she may insist vehemently that what she/he said is the same as what the modeling adult has said. It is of course difficult to know whether the child is referring to pronunciation or to content in such assertions; metalinguistic conversations with two-year-olds tend to be unsatisfactory (Brown & Bellugi, 1964; in Brown, 1970, p. 79).)

The third type of unexpected variation is again a bimodal variation brought about by rule conflict, but this time it is not a passing unstable phase marking the cusp-point of change. Instead, it seems to reflect the co-existence of competing rules which may arise and decay at about the same time (Menn, 1973). We will refer to this as rule-coexistence variation when it is necessary to distinguish this type of rule-conflict variation from transition variation.

A fourth interesting kind of variation, which we will call floundering, can be described as wide fluctuation in the production of a particular model phone or string of phones under phonologically stable conditions. An example is Daniel Menn's 'peach' attempts, [itš] [dits] [pipš] [gik] [nitš] etc. (Menn 1973). This kind of variation I have interpreted as being what happens when a child has no well-formed rule for dealing with a particular string of phones, that is, where the model word does not meet the structural description of any of the child's rules, and where the outputs look like what would happen if one or several features of the model word were changed so that it could be an input to the child's rules. Conceptually, floundering is quite distinct from backgrounding; floundering is the result of trying to use rules that don't quite apply, while backgrounding occurs when the child's output is produced with less reliance on practiced rules and more attention to pronunciation as a task. The parallel distinction can be made in adult second-language learning. Suppose we have an American trying to pronounce a hypothetical word /ndaŋa/, containing the unEnglish cluster /#nd/ and the morphologically controlled medial /ŋ/. Suppose our speaker is able to get each of these difficult items correct when thinking about it, but that s/he otherwise reverts to initial /#end/ or /#d/ and to medial /ŋg/. The variation between /ŋ/ and /ŋg/ (and also between /#nd/ and either of the two wrong pronunciations) is controlled by the amount of attention that it gets: this is

backgrounding. On the other hand, the variation between /#end/ and /#d/ for /#nd/ is floundering: it is a random choice among sounds which have a close resemblance to the difficult target.

Finally, some young children show lexically controlled variation. Here, certain words show great variation in the production of some or all their sounds while other words that have similar adult models show much less variation. Jacob (Menn 1976) had a much greater variability for the /æwn/ sequence in 'down' than for the same target in 'around'. This also has parallel at the margins of adult phonology: consider for example the great variety of sounds permissible (as expressive variants) for the 'phoneme' /o/ in the word 'no'. This variety is not found in renditions of the same phoneme in the word 'know'.

We have named seven types of variation of special interest to child phonology. Now, by a 'model' of a phonological system, I mean a flow chart which specifies roughly what information is stored, what is used in real time, and how the different pieces are brought together to specify the articulatory instructions needed to produce a word. How can these seven types of variation be represented in such a model?

The most important capability to be added to extant models actually is, I think, one that has not been explicitly mentioned so far, since it manifests itself indirectly. Child phonology models almost all represent the steady state: the rule or word is established. These models need new apparatus to simulate what happens when a new word is being tried or a new rule is being formed, for practiced behavior is very different from novel behavior. This familiar-novel distinction seems to be related to the distinction that we have already invoked between monitored and automatic behavior, but they are not the same. To deal with both novelty and attention, models will have to allow more than one route from adult word to child word. We could say that one route, the one used most frequently, would represent automatic, over-learned behavior, and other routes would correspond to the special cases when at least part of a word is not being produced under automatic control. We can make this more explicit by considering an available child-phonology model.

Suppose we use a two-lexicon model similar to the one in Kiparsky and Menn (1977), concerning ourselves with the part of

it that would run: adult form + (perceptual strategies) + phonetic representations perceived by child = input lexicon + (reduction rules) + encoded articulatory representations = output lexicon + (motor routines) + child's output form. The input lexical entry represents the child's encoding of his/her percept of the adult word, the output lexical entry represents an encoding of articulatory instruction, and the reduction rules relate the two lexicons.<sup>1</sup> We can modify such a model to allow for non-automatic speech production by adding routes from the input lexicon (percept of model word) to the output side (pronunciation) that bypass the output lexicon and some of the rules that lead into and out of it. This would represent an attempt to give a spontaneous rendition of a known word without most of the automatic apparatus, and might represent what goes on during word-practice. To represent imitation, we would also add routes from some point(s) among the perceptual processing routines that would bypass both lexicons and feed into some points among the articulatory routines.

The variation in the points of beginning and ending of these bypasses would reflect the degree to which established perceptual and articulatory routines were employed in the utterance. (Presumably, the more that one monitors, the more habits of perception and production can be overcome.)

It seems, then, that some aspects of transition (rule change), backgrounding, and imitation-effect variation can be modeled by the addition of these new processing 'routes' to a K & M-type model. It turns out only a few more entities are required to adapt this model or its descendants to represent the other four types of variation that we have discussed.

Coexistence variation can be simulated by letting both of the competing reduction rules operate on each applicable input lexical item, thus generating two forms in the output lexicon corresponding to each of those input forms. Either of those forms could be translated into output any time the child said the word. If the probabilities that the two forms both occur are not equal, some notion of the 'strength' of a lexical entry must also be added, so that one could say that the stronger entry is the

(1) The recent revision of the K & M model presented in Menn 1977 would allow a clearer formulation of some of the following discussion, but occasions no major differences.

one produced more frequently.

Transition variation would also be represented by having two output lexical entries, one generated by the older rule and one by the new rule. As we have implied, we can model the loss of a rule by removing it from the set of production rules. This will 'disconnect' some output lexical entries from their input lexical entries. (Transition variation would thus not be rule competition, as we stated above, so much as competition between two output lexical entries.) Since new rules normally spread to older words, we might hypothesize that the 'disconnected' output lexical entries lose strength and fade away. However, we know that some lexical entries which clearly do not have live support, such as phonological idioms and fossils (words which inexplicably resist rule changes), do not fade in the usual way but remain vigorous for long periods. If the 'fading' notion is used, we require special apparatus to handle phonological idioms and fossils. Several have been proposed (see Macken 1978) but we cannot pursue that topic here.

Local scatter does not involve lexical entries at all, but has to do with the lowest output processing levels: we shall assume that it occurs when articulatory instructions for a phone are executed with more tolerance than they would be by an adult.

Lexically controlled variation, on the other hand, requires, obviously, a special entry in the output lexicon just as phonological idioms do, and in addition this entry must specify special articulatory instructions rather than the general output routines or in addition to them.

The remaining form of variation that we have discussed is floundering. The basic situation in floundering seems to be a rule-input that is ill-formed. The proper analysis of a given case, however, may depend on the whole rule-structure, because there are several ways that this could happen in the present type of model. There are two relevant loci: the input lexical entry could fail to meet the structural description of necessary reduction rules, or the output lexical entry could fail to be of the proper form for the articulatory instructions to handle. In addition either case of ill-formedness might be better modeled by overspecification, underspecification, or some other type of malformation. Further elaboration of the psychological interpre-

tation of this or similar models of child phonology will be required in order to make a principled choice among these alternatives.

To conclude: certain types of variation are intimately and essentially involved with learning to pronounce. As we build richer models of child phonology, we can incorporate them without undue difficulty. Regardless of how easily we can draw new lines and little boxes, however, one problem about transition and variation remains very difficult. How does a new linguistic behavior cease to be effortful and become automatic?

#### References

- Brown, R., and U. Bellugi (1964): "Three processes in the child's acquisition of syntax", in Psycholinguistics, R. Brown, Glencoe, Ill: The Free Press.
- Ferguson, C.A. and C.B. Farwell (1975): "Words and sounds in early language acquisition", Lg 51:419-439.
- Kiparsky, P. and L. Menn (1977): "On the acquisition of phonology", in Language learning and thought, J. Macnamara (ed.), New York: Academic Press.
- Macken, M.A. (1978): "The child's lexical representation: the puzzle-puddle-pickle evidence" (ms.), Stanford University Linguistics Department.
- Menn, L. (1971): "Phonotactic rules in beginning speech", Lingua 26:225-251.
- \_\_\_\_\_ (1973): "On the origin and growth of phonological and syntactic rules", Papers from the Ninth Regional Meeting of the Chicago Linguistic Society, 378-385.
- \_\_\_\_\_ (1976): "Pattern, control, and contrast in beginning speech: a case study in the development of word form and word function", U. Illinois doctoral dissert., to be circulated by the Indiana University Linguistics Club.
- \_\_\_\_\_ (1977 and to appear): "Phonological units in beginning speech", in Syllables and segments, A. Bell and J. Hooper (eds.), Amsterdam: North Holland Publishing Co.

## SPEECH SOUND CATEGORIZATION BY CHILDREN

Paula Menyuk, Applied Psycholinguistics, Boston University,  
Boston, Mass., USA

Clearly acquisition of the structural properties of language takes place via a process of segmenting and then categorizing the segments of the language heard into units which can be used to comprehend and generate unique utterances. Equally clearly, the first aspect of language that is used by the hearing adult and by the infant to segment and categorize utterances is the acoustic speech signal. Unlike the adult, however, who has already determined what the "appropriate units" are and can by-pass much of the surface structure of the utterance, the infant must rely heavily on the signal to come to conclusions about appropriate segmentations. She must also rely heavily on contextual cues to relate the segments of the signal to objects and events in the environment. Despite the fact that common sense tells us that the above must be the case, we are, at the present time, still unclear about what these segments are and what the bases for categorization of segments are either initially or over time as the child matures. Indeed, controversies still exist in the literature on this issue for the adult as well as for the child. In this paper varying hypotheses concerning the nature of and bases for speech sound categorization by the child and its role in language acquisition will be examined in light of theories on adult language processing and the data on the speech processing behavior of the infant and young child.

Theoretical Descriptions of Processing

The segmentation of continuous speech has been described by linguists as being hierarchical and nested. That is, a message can be characterized as a type of speech act. The message can contain a sentence or sentences and these contain phrases which are made up of morphemes. Morphemes are composed of syllables which are made up of speech sound segments each of which represents a bundle of features. If this were a psychologically real description of language processing as well as of elements of language then the listener would determine categories of segments in a sequence with each lower step of the sequence dependent on the immediately higher step since higher steps indicate units of analysis. Speech sound segment categorization would take place at the end of the sequence

and require resolution of the bundle of features. Varying descriptions based on this hierarchical model have been labelled "analysis by synthesis" (Cooper, 1972). It might be logically argued from this model that since speech sound categorization or identification is comparatively late in the sequence of on-line processing then it must also be late in the sequence of acquisition of the structural properties of the language.

The above model of language processing has been deemed inadequate in accounting for either real-time processing of speech by the adult or for the observed sequence of acquisition of structural properties by the child. Since earliest utterances are sequences of speech sounds marked prosodically and the infant does not appear to understand anything more about utterances than their affective intent, it cannot account for behavior during the early babbling period. The model does not account for subsequent language behavior since even then the child does not evidence any knowledge of any of the postulated higher categories (i.e. sentence, phrase and morpheme). An alternative description of both processing and the sequence of acquisition is a bottom-up model or "synthesis by analysis". With this model speech sounds are differentiated and categorized, then grouped into higher level categories in a sequential manner during speech processing. In acquisition, speech sounds are differentiated and categorized by a process of imitation and sound approximation which is rewarded. These sounds are then composed into words by the same process and by associating phonological sequences with objects and events. Larger units of an utterance, phrases and sentences, are composed by putting together smaller units via a chaining process (Staats, 1971). This description suggests that the earliest analysis in processing and the earliest structural acquisition are segmental speech categories although the nature of these categories is not defined in the model.

Not only is there a substantial amount of evidence to indicate that this model does not adequately describe adult language processing (Fodor et al., 1974), but, also, it is difficult to see how the analysis of speech sounds one by one in a sequence can lead to decisions about where crucial boundaries lie and, thus, to a determination of meaning. For these same reasons a synthesis by analysis model seems inadequate in accounting for language acquisition. Although there is evidence that in early care-giver-child



communicative interaction, segments and boundaries are made much more salient than they are in adult-adult communication (Newport, 1976), there is no evidence that the child, in the process of acquisition, adds sequentially to segments by chaining bits together or that the child merely imitates input structures. On the contrary, the child appears to be only able to attend to and generate certain aspects of utterances at certain periods of development regardless of input structure and these aspects are not sequential bits of adult utterances.

Still another description suggests that perception and generation of connected speech is a parallel process; i.e. one involving all components of the language simultaneously. In the process chunks of the message, probably phrases, are subjected to analysis and rough estimates are made of the phonological composition of the morphemes in the phrase to corroborate hypotheses about the meaning of the phrase and then other phrases if more than one is contained in the utterance. An exact representation of the phrase can be kept in mind until analysis is completed so that needed corrections on this estimate can be made (Garrod and Trabasso, 1973). The child, during the process of acquisition, would analyze the data in the same fashion. The distinctions between the child and the adult are in the amount of information chunked for analysis, the much heavier reliance on the part of the child on contextual cues for analysis and the process of chunking itself since segmentation strategies would change as more structural knowledge of the language is acquired (Menyuk, 1977, Chap. 5). For example, an early chunking strategy might be to ignore everything in the signal except those sequences that signal main relations of actor and action or action and object. Components of the relation would be grossly analyzed for lexical look-up. However, analysis of the phonological segments per se would not be needed for comprehension. Since the parallel processing requires analysis of segments only when correction of rough estimates is required it might, again, be logically argued that speech sound segment categorizations would be later acquisitions than morpheme categorizations.

The above are theoretical descriptions of adult language processing and theoretical descriptions of the process of language acquisition. In conjunction with these are descriptions which are concerned with phonological acquisition only. This acquisition has

been described as a process of first discriminating between the speech sounds of the language, then categorizing these distinctions in terms of articulatory gestures. These discriminations are based on distinctive feature differences between speech segments. Early distinctions are determined by feature detectors that are pre-programmed in the auditory system of the human infant (Eimas, 1974). These might be termed primary features. Finer distinctions are then made both in perception and production and are probably affected by particular language experience. However, given the universality of the speech processing abilities of normal infants, both perceptually and productively, there is, to some extent, universality in the sequence in which distinctions are made. This universality is confounded by the particular data the child is confronted with; i.e. the language of the child's community and even family. Thus, the universal order is modified by the perceptual and productive problems a particular language poses for the child and by the interactive styles, lexical selections, etc. of a particular family. Individual differences become more marked when standard lexical items in a particular language begin to be used.

Data on Early Speech Processing

On the face of it there appears to be a logical gap between theories of language processing and of language acquisition and theories of the development of the phonological system. The latter suggest very fine analysis of the signal on the segmental level in terms of distinctive features, whereas the former suggest rather gross analyses dependent on higher level categories. There are also large differences between the findings of studies carried out at different periods of early speech processing behavior. One of the primary reasons for these gaps between theories of language and speech acquisition and between the findings of studies of speech processing and the conclusions drawn from them may be not maintaining a clear distinction between what the infant and child can do and what they ordinarily do; i.e. a capacity versus performance distinction. The data collected thus far on early speech processing indicate that the very young infant (1 to 4 months) as well as the very young child (under two years) can discriminate between speech sound segments that vary in terms of a single distinctive feature. There also appears to be a hierarchy in the features that can be distinguished both perceptually and productively. Thus, during



the cooing and babbling periods some features appear to be more perceptually salient than others and this also appears to be the case when the task is distinction of minimal pair nonsense syllables. Similarly, segments containing certain features are realized before segments containing other features in babbled utterances and then in morpheme production. However, there is not an exact correlation between the order of perceptual and productive distinctions made, and individual differences in the exact sequence of features and segments distinguished can be observed.

What seems to be suggested by these data is that distinctions can be made on the basis of distinctive features by the infant and young child if the question is put to them in a way in which these distinctions are made clear; i.e. in a small enough context that is non-distracting such as nonsense-syllables. Also, the response required in the task must be part of the children's behavioral repertoire. For example, they must have sufficient memory to recall the stimuli presented. Finally, there are some features that can be distinguished before others. However, discrimination between features does not imply that categorization of segments has taken place in terms of bundles of features nor does the capacity to discriminate between features imply that this is what children do when they listen to speech and attempt to match articulatory outputs to stored representations. Indeed, all the data indicate that during the babbling and early lexical acquisition periods distinctive feature differences are not actively employed in determining meaning of utterances or in generating utterances.

During the babbling period perceptual processing of continuous speech seems to be primarily based on the supra-segmental aspects of the speech signal and contextual cues. Some time toward the end of this period recognition of a small set of lexical items is observed and still later production of word approximations begins. The lexicon of the child at this time is quite small. It is entirely reasonable to suppose that both lexical recognition and generation are based on syllabic representations of morphemes. In other words, speech processing is taking place on the basis of the morpheme and this may be the minimal unit for categorization of speech information. The meaning of a phonological sequence, its gestalt phonological representation as a syllable or reduplicated syllables, supra-segmental features of intonation and contextual cues appear

to be all that is needed or used to comprehend or generate utterances during this time (Menyuk and Menn, in press).

Again, this is what children appear to do in on-line processing of speech during these early periods of development, although, at this time and long before, they are capable of discriminating between speech sounds on the basis of feature distinctions. As the lexicon grows and as structural knowledge increases constraints on memory probably make segmental differentiation and categorization necessary. When this occurs an available competence is actively employed. However, segmental differentiation and categorization may be needed only rarely to comprehend continuous speech. Thus, although the ability may be increasingly used at later periods of development it still may be used infrequently. Research shows that even 3 and 4 year-old children first use morpheme information to differentiate between phonological sequences and only use segmental information with some exertion when morpheme information is unavailable; i.e., with nonsense syllables or unknown words. At present, little is known about when reference to segmental information is used without marked exertion. Such ability is, of course, required in learning to read alphabetic text. One would assume that this ability develops gradually and that there would be individual differences or group variations due to language experience in the ages at which this ability manifests itself (Savin, 1972).

#### Conclusions

The theoretical description of the processing of language which appears to most adequately describe the sequence of acquisition of the structural properties of the language and to best fit the data on infants and young children's speech processing is one of parallel analysis of chunks of continuous speech. Initially the chunks the child can process are short in duration, linear in arrangement and involve primarily surface structure information. Reference is made to gestalt representations of surface acoustic information to derive meanings. Thus, the analyses are quite gross. As the child matures the chunks that can be processed simultaneously at all levels (semantic, syntactic and phonological) increase in duration and, as structural knowledge grows, recursiveness within chunks can be processed and the analysis becomes more detailed or differentiated. The speech signal must be held in mind and represented to allow analysis using whatever structural knowledge is available. It has been suggested that this representation

or categorization of speech is initially acoustic images of morphemes and/or syllables and only later in terms of segments and features of segments. This appears to be the case even though the infant is capable of discriminating between minimally different acoustic features. In summary, the model that appears to be most descriptively adequate is not a "top-down" or "bottom-up" model but, rather an "outside-in" model (Menyuk, 1977).

#### References

- Cooper, F. (1972): "How language is conveyed by speech", in Language by ear and by eye, J. Kavanagh and I. Mattingly (eds.), 25-46, Cambridge, Massachusetts: MIT Press.
- Eimas, P. (1974): "Linguistic processing of speech by young infants", in Language perspectives: acquisition, retardation and intervention, R. Schiefelbusch and L. Lloyd (eds.), 55-74, Baltimore: University Park Press.
- Fodor, J., T. Bever and M. Garrett (1974): Psychology of Language, New York: McGraw-Hill.
- Garrod, S. and T. Trabasso (1973): "A dual memory information processing interpretation of sentence comprehension", JVLVB 2, 155-167.
- Menyuk, P. (1977): Language and maturation, Cambridge, Massachusetts: MIT Press.
- Menyuk, P. and L. Menn (In press): "Early strategies for the perception and production of words and sounds", in P. Fletcher and M. Garman (eds.), Cambridge, England: Cambridge University Press.
- Newport, E. (1976): "Motherese: the speech of mothers to young children", in Cognitive theory: Vol. II, N. Castellan, D. Pisoni and G. Potts (eds.), Hillside, New Jersey: Lawrence Erlbaum Assoc.
- Savin, H. (1972): "What the child knows about speech when he starts to learn to read", in Language by ear and by eye, J. Kavanagh and I. Mattingly (eds.), 319-326, Cambridge, Massachusetts: MIT Press.
- Staats, A. (1971): "Linguistic-mentalistic theory versus an explanatory S-R learning theory of language development", in Ontogenesis of grammar, D. Slobin (ed.), 103-152, New York: Academic Press.

SOCIAL FACTORS IN SOUND CHANGE: Summary of Moderator's Introduction  
Einar Haugen, Department of Linguistics, Harvard University, Cambridge, MA 02138, U.S.A.

The papers offered in this symposium may be divided into "theoretical" and "empirical", even though of course both types of research are represented in all. The papers by Birnbaum, Fónagy, and Malmberg are primarily theoretical, Brink/Lund, Labov, and Peng primarily empirical.

Birnbaum offers for discussion a model of linguistic change originated by Henning Andersen, in which the key word is "abduction", especially applicable to the process of linguistic decoding. Birnbaum is critical of certain aspects of this model, especially its implication that a speech community may be homogeneous or consist of neatly separable generations.

Fónagy is concerned with the idea offered by some that intonation is a non-arbitrary, naturally motivated phenomenon. To disprove this he offers samples from French and Hungarian of how intonations can change their signification over time and become arbitrary expressions associated with particular social groups.

Malmberg takes as his starting point his own earlier studies of the Parisian vowel system, in which he found an "état de langue" which included two systems, a "maximum" and a "minimum" system of vowels between which the speaker could choose. The "minimum" system represented a simplification, which Malmberg attributes primarily to "peripheral" learners of the language, whether they be socially or geographically marginal, i.e. lower class or colonial, the latter exemplified by Spanish in the Americas.

Brink and Lund (here Brink/Lund) have completed a massive study of Copenhagen speech from 1840 to 1955, based on the recorded voices of speakers born between these dates. Their researches have uncovered some 60 phonetic changes (which they call "sound laws") that characterize this period and permit them to classify their speakers into two groups, according to whether they speak "high" or "low" Copenhagen.

Labov's paper sums up some of the conclusions at which he has arrived on the basis of his classic studies of Martha's Vineyard in Massachusetts, the Lower East Side of New York City, and the city of Philadelphia. He has been a pioneer in developing a technique of selecting "social markers" which permit him to place speakers rather accurately on the socioeconomic scale.

Finally, Peng presents a summary of his studies of the linguistic changes in the city of Tsuruoka in Japan, data gathered by his colleague Nomoto in the years 1950 and 1971, in many cases from the same informants. Out of this material he has drawn conclusions that reduce the time span within which one can observe linguistic change even more drastically than Labov: he contends that it is possible to identify linguistic change within a single generation.

Each of these papers brings something to the elucidation of a problem that has baffled linguists ever since the regularity of sound change was firmly established early in the nineteenth century. The causes of sound change were vainly sought in everything from climate to human physiology. Until recently linguists were convinced that change was so slow that it was inaccessible to direct observation. Diachronic linguistics became the study of the past, historical and even paleontological, while a synchronic linguistics sprang up which was based on assumptions of heuristic stability and uniformity, as language might wishfully appear to the prescriptive grammarian.

The Prague School declared that the ideal standard language should possess both stability and elasticity, i.e. it should be flexible enough to change and yet conservative enough to seem unchanging. They did not realize that this paradox could and must apply to every variety of human language; its latest synonym is Labov's expression in describing his concept of language: "orderly heterogeneity". He opposes this to Chomsky's "ideal homogeneity", but in fact his variable rules are a formalization of the concept of "elasticity", while categorial rules reflect "stability". Questions have been raised about the statistical nature of variable rules: how can a speaker know that he is going to use one sound 66% of the time and another the remaining third?

Part of the answer comes from the painstaking analysis by Brink and Lund of the recorded materials from Copenhagen. They have offered no statistics, but in their big book (unfortunately available only in Danish) they have traced from decade to decade how certain changes arose, how speakers vacillated from one to the other form, and how new generations resolved the conflict by choosing one or the other of the alternatives. It is clear that the concept of "choice" with which Malmberg operates has been at work, but it is not clear that it has been a choice between two or more coherent levels of speaking. Even with the masses of data now being accumu-

lated in such studies, including Labov's and Peng's, we are far from knowing why these choices are made, either individually or collectively. Such a study would be an infinite regression going far beyond the realm of linguists' competence, especially if the goal were to construct some kind of predictive model that would tell us what kind of changes the future will bring. Brink/Lund's material shows clearly that at any given point in time there is a great deal of unstructured heterogeneity, vacillation which may either lead to innovation or regression.

Our contributors differ sharply on certain crucial aspects of the problem. Brink/Lund flatly assert that regular "sound changes occur between generations (in our opinion innovations come from children, who - under mutual influence - retain while growing up a few of their originally many deviations from the adult language)". They refer to recordings of the same persons from 30 to 50 years apart in which one could detect virtually no change. Against this Peng claims that the changes passed on are those that young people have developed up to the age of 35, when they communicate them to their children. Against these extreme views we may place Birnbaum's judicious remark that there is a "continuous pattern-setting effect of parents on children, teachers on students, leaders on followers, older on younger playmates and fellow workers, more prestigious on less prestigious...".

There is also some difference of opinion on the role played by social classes and other groups in the activation of change. Labov has found that the upper working or lower middle class leads in changing, while Brink/Lund hold that in general the lower classes of Copenhagen have been in the lead, as being the majority, if not the most prestigious socially. The difference may be more terminological than real, for it is hard to compare the finely graded scale of socioeconomic status developed by Labov with Brink/Lund's linguistic division of the entire population of Copenhagen into two groups, the H-speakers and the L-speakers. On one point all are agreed: that women have more H-features than L-features, though Brink/Lund will not grant that there is a special female sexlect.

Both Peng and Malmberg emphasize that it is not language that changes, but people who change language. This is clear enough when we speak of the adoption of new words or the learning of new dialects and languages, but for phonology the functioning is so automatized and deeply embedded in the subconscious that it has been

difficult to find any clear social causes for specific changes, e.g. Umlaut or the Germanic consonant shift.

I would suggest that we do know a good deal about the causes of sound change, but we have made little progress in predicting its results. But at least we now have techniques and instruments that enable us to catch it on the wing and study it while it is going on. We still have a long way to go before we can learn to control it, if we should ever wish to do so. In this respect we are no worse off than any other social science.

ONGOING SOUND CHANGE AND THE ABDUCTIVE MODEL: SOME SOCIAL  
CONSTRAINTS AND IMPLICATIONS

Henrik Birnbaum, University of California, Los Angeles, USA

Underlying the present discussion of some aspects of sound change is the notion that language not only, as energeia, (or, explicitly, as a set of largely automatized processes definable in more or less accurately phrased rules), is susceptible to formal analysis of some degree of descriptive adequacy and explanatory power but that, in addition, it can be conceived of as an inherent and integral part of human thought and imagination. Adopting the latter point of view, language can be said to form a conceptualized (verbalized) mirror image of mental activities (cf. the notion of language as the primary modeling system, elaborated in Soviet semiotics). The former approach, concerned with building models of linguistic structure (or parts thereof), views language as a - particularly sophisticated - semiotic subsystem (operating within the parameters set by its specific neurophysiological premises) and strives to explain its functioning in this capacity. The other kind of inquiry into the nature of verbal communication places the chief emphasis on language as a cultural manifestation of the human mind (in the sense of Geisteswissenschaft) and seeks to understand its performance in society. The former approach may be termed generative (in the broadest meaning), the latter hermeneutic. Both, if applied pragmatically and without any ad hoc constraints, have a sociolinguistic dimension.

It is a fairly common view that sound change takes place gradually in a series of minimal, barely noticeable adjustments and modifications at the phonetic (subphonemic) level and that it is only at the functional or semantically distinctive (phonemic) level of sound production and, in particular, perception that the impression of abrupt sound change obtains.

Some years ago, Andersen (1973), while critical of 'standard' TG phonology but adopting a broadly generative approach to linguistic inquiry in terms of positing specific speaker/hearer 'grammars', i.e., sets of rules generating acceptable sound sequences (utterances), proposed an intriguing model of phonological change. In addition to induction and deduction, he introduced, following Peirce, a third mode of inference termed abduc-

tion. Applying deduction and abduction specifically to sound change, Andersen (1973, 777, fn. 13) points to the "unique role of abduction ... vis-à-vis the other modes of inference, which merely test what has been arrived at by abduction" and suggests that "one can evidently describe the process of encoding as essentially deductive, and that of decoding as abductive". In closing, he submits (1973, 791) that while early structuralism (Jakobson) "could insist only that every phonetic innovation be interpreted in terms of the system that undergoes it ..., it is [now] possible to interpret every phonological innovation - abductive or deductive - in terms of the system that gives rise to it".

In a subsequent paper, Andersen (1974, esp. 25-6, 41), in discussing and summarizing his typologies of innovation in the content and expression systems of language, distinguishes between adaptive and evolutive innovations, with the former subclassified, on the expression plane, into remedial and contact innovations; the evolutive innovations are subdivided into deductive and abductive, with the abductive innovations of the expression plane further specified as pertaining either to the phonemic system (a) feature valuation, b) segmentation, c) ranking), or to pronunciation rules. In a more recent study, with his theoretical reasoning again firmly grounded in Slavic diachronic and dialectal data, Andersen (1978, section 4.2) arrives at the conclusion that we must "acknowledge that conceptual factors take precedence over perceptual or articulatory ones in determining how a phonological system may be changed as it is transmitted from generation to generation ... and recognize that it is the structuring principle of linguistic form - the fact that the speech signal must be segmented, that distinctive features are binary, and that they must be ranked - and not the articulatory or acoustic or perceptual substance that shape its historical development. We are led to conclude that the ultimate source of dialect divergence - and of linguistic change in general - is the process of language acquisition, in which the speakers of a language impose form on the fluctuating and amorphous substance of speech." Novel and incisive though these formulations are, they not only allude to Jakobson's views about DF analysis and language acquisition, but in their reference to form and substance, content and expression also echo some of the basic tenets of glossematic theory. Yet, essential-

ly, the abductive model of sound change, pertinent, above all, to the decoding process, is of course Andersen's, at least as consistently formulated by him and solidly underpinned by theoretical considerations. The model implies that the output of 'grammar 1' serves as the input to 'grammar 2' which in turn yields a reinterpreted 'output 2', slightly, yet significantly different from 'output 1' (1 and 2 here symbolizing successive generations); cf. esp. Andersen (1973), 767 and 778, figs. 1 and 2.

It should be noted, however, that observations and inferences of a similar kind have been made with regard to phonological change also prior to Andersen's sketching of his model of abductive innovation in phonology, as well as after the appearance of his first, seminal paper on the subject. As an example of the latter - arrived at independently, it seems - may be quoted some remarks made by Hetzron in discussing two principles of reconstruction in genetic linguistics. Thus, Hetzron (1976, 96) writes: "In diachrony ... what is transmitted from generation to generation is not the structure, but a set of data which is analyzed by the child acquiring the language so that he could establish a structure for his own use. Language change is precisely justified by the fact that a subsequent generation may analyze the facts perceived by learning the language from the older generation, and this may eventually require some adjustment in the facts, some modification of the perceivable data". To be sure, Hetzron's formulation is less precise than Andersen's in addition to being couched in traditional structuralist ('taxonomic') rather than in broadly generative terms. But in essence, this is in line with Andersen's more elaborate and tightly argued model of phonological innovation.<sup>1</sup>

When stating his premises, Andersen (1973, 767) wrote: "What is needed is a model of phonological change which recognizes, on the one hand, that the verbal output of any speaker is determined by the grammar he has internalized, and on the other, that any speaker's internalized grammar is determined by the verbal output from which it has been inferred." And he qualified

(1) For an earlier comment on the similarity of Andersen's and Hetzron's reasoning and a first criticism of a shortcoming they, in my opinion, share, see Birnbaum (1977), 28-30.

his theoretical framework by adding the crucial requirement: "The model that is needed must show how phonological innovations can arise in a homogeneous speech community ..." While the broadly generative (and logic) premise sketched seems most useful indeed, the formulation of the sociolinguistic condition is somewhat questionable (his reference to Labov's definition notwithstanding). What, in fact, is a homogeneous speech community? And what exactly is meant when Andersen (like Hetzron) speaks about the transmitting of a phonological system (or a set of data) from generation to generation? As I had an opportunity to caution (Birnbaum, 1977, 30): "... the transmission of a linguistic system or subsystem (or a grammar or grammatical component generating this system or subsystem) from one generation of speakers to the next must not be conceived of in all too rigid, mechanistic terms since the distinction of successive generations in any real speech community is never very clear-cut and easily ascertainable." Put differently, even though sound change in reality — on the phonetic level, accessible to physical scrutiny and measurement — occurs gradually and it is only on the more abstract phonemic level that one sound, at some point, simply replaces another, it is nonetheless a fact that, given the passage of time, an actual sound shift (e.g., *e* > *o*, *ou* > *u*; *d* > *t*, *k* > *č*) is ascertainable also at the phonetic level. How do such phonological changes come about? Surely not as a result of any simultaneous gradual adaptation by each entire membership of a number of clearly definable consecutive generations. Obviously, a real speech community is never truly homogeneous, nor does it consist of a limited set of neatly separable generations.

Considering the interpenetration of synchrony and diachrony — in phonology, ongoing sound change — it would seem more realistic not to posit a limited set of coexistent generations at any given time (as is implied in Andersen's abductive model as well as in Hetzron's informal reasoning) but rather to assume the continuous pattern-setting effect of parents on children, teachers on students, leaders on followers, older on younger playmates and fellow workers, more prestigious on less prestigious population groups, etc., all interacting at various ages and stages of their development. While such a view of society and language does not vitiate the validity of Andersen's abductive model of sound

change altogether, it certainly makes his scheme more problematic; also, given these complicating factors, his technique for describing, analyzing, and explaining actual phonological innovation is in need of further refinement.

Here one more point should be briefly discussed. It has become customary to attribute great significance to the process of acquiring language, i.e., the mastering of one's native tongue in early childhood, also when it comes to explaining certain basic facets of sound change. (The partial or complete acquisition of a foreign language presents analogous but also additional problems.) Andersen's abductive model, in this respect influenced by Jakobson's work on child language, is but one example of this conception. However, it seems worth considering whether, precisely as regards modifying one's pronunciation habits, i.e., introducing incipient or, occasionally, even full-fledged phonological innovations, it is actually in early childhood (say, before the completion of the fifth year) that the definitive articulatory profile of a person is usually formed and stabilized. Rather, I would submit, that is the age when growing-up speakers, by imitating their elders, attain the same or nearly same pronunciation as their models. True, in the process they may very well, by 'misreading' (i.e., slightly incorrectly perceiving) the phonetic output of 'grammar 1', internalize, initially, at least, a somewhat deviant 'grammar 2' (or, rather, its phonological component) producing — following Andersen's reasoning — a phonetic 'output 2' not fully identical with 'output 1' of their model. Yet, very often (if not as a rule) most of the misperceived pronunciation is subsequently noticed and rectified except, perhaps, where the resulting differences in pronunciation are so minimal as to be considered insignificant even by the maturing child; it is only their cumulative effect over a longer period of time that ultimately may give rise to a genuine sound change. However, it appears that attitudes at a somewhat older age, especially in the teens, may more directly, noticeably, and lastingly affect pronunciation habits and cause partial or even full sound shifts (or, rather, sound substitutions) to occur within one generation. I am referring here to the fashionable pronunciation or talking fads which, particularly in our day and age, so markedly leave their imprint on the speech habits of the teenage generation. It



is my impression, based on observations from several languages, that the modification of the articulatory manners and preferences affecting these young people are more radical, since they are deliberate, than are the difficulties in imitation and pronunciation adjustment encountered in early childhood. If Andersen's abductive model of phonological innovation is to be applicable also to currently observable sound change – and not only to interpreting and elucidating instances of historically attested or reconstructed phonological shifts – these sociolinguistic and psycholinguistic considerations will somehow have to be accounted for in his model.

Viewing sound change primarily as a sociolinguistic phenomenon, best studied while in progress, it must be said – with all due respect to Labov's 'integrated' explanation<sup>2</sup> – that we are still far from genuinely and fully grasping its causes. So far, there has not been much more than a general realization of the permanent and highly creative interplay between, on the one hand, language's striving for economizing (ultimately tending toward ellipsis while preserving a measure of redundancy as a safety valve to ensure comprehension and information transfer; cf. Martinet 1955) and, on the other, its making for diversity of expression to distinguish among even the finest shades of meaning. Though sound, at the phonemic level, does not by itself carry, but merely distinguishes meaning, it and its modification are crucially affected by this dialectic tension characteristic of language as a semiotic system.

(2) The study of ongoing sound change viewed in its social setting has in America been pursued, in particular, by Labov; cf. esp. Labov (1963), (1966), (1970), (1972), (1973); and Labov et al. (1968), (1972); for a brief assessment of Labov (1973), see, e.g., Birnbaum (1975), 284-6. Of more recent work by scholars with other ideas, see, e.g., Bailey (1973), Peng (1976), and Itkonen (1977).

#### References

- Andersen, H. (1973): "Abductive and Deductive Change", *Lg.* 49, 765-793.
- Andersen, H. (1974): "Towards a Typology of Change: Bifurcating Changes and Binary Relations", in: *Proceedings of the First International Conference on Historical Linguistics* (J. M. Anderson & C. Jones, eds.), II, Amsterdam & Oxford: North-Holland, 17-60.
- Andersen, H. (1978): "Perceptual and Conceptual Factors in Abductive Innovations", in: *Recent Developments in Historical Phonology* (J. Fisiak, ed.), The Hague: Mouton [in press].
- Bailey, C.-J. N. (1973): *Variation and Linguistic Theory*, Arlington, Va.: Center for Applied Linguistics.
- Birnbaum, H. (1975): "Typological, Genetic, and Areal Linguistics: An Assessment of the State of the Art in the 1970s", *FoundLg* 13, 267-291.
- Birnbaum, H. (1977): *Linguistic Reconstruction: Its Potentials and Limitations in New Perspective*, Washington, D. C.: Institute for the Study of Man (*The Journal of Indo-European Studies*, Monograph No. 2).
- Hetzron, R. (1976): "Two Principles of Genetic Reconstruction", *Lingua* 38, 89-108.
- Itkonen, E. (1977): "The Relation Between Grammar and Sociolinguistics", *Forum Linguisticum* I:3, 238-254.
- Labov, W. (1963): "The Social Motivation of a Sound Change", *Word* 19, 273-309.
- Labov, W. (1966): *The Social Stratification of English in New York City*, Washington, D. C.: Center for Applied Linguistics.
- Labov, W. (1970): "The Study of Language in Its Social Context", *Studium Generale* 23, 30-87.
- Labov, W. (1972): *Sociolinguistic Patterns*, Philadelphia: University of Pennsylvania Press.
- Labov, W. (1973): "The Social Setting of Linguistic Change", in: *Current Trends in Linguistics* (T. A. Sebeok, ed.), 11, The Hague & Paris: Mouton, 195-251.
- Labov, W. et al. (1968): "Empirical Foundations for a Theory of Language Change" (with U. Weinreich & M. I. Herzog), in: *Directions for Historical Linguistics: A Symposium* (W. P. Lehmann & Y. Malkiel, eds.), Austin & London: University of Texas Press, 95-188.
- Labov W. et al. (1972): *A Quantitative Study of Sound Change in Progress*, 2 vols. (with M. Yaeger & R. Steiner), Philadelphia, Pa.: U. S. Regional Survey (NSF GS-3287).
- Martinet, A. (1955): *Économie des changements phonétiques*, Berne: Francke.
- Peng, F. C. C. (1976): "A New Explanation of Language Change: The Sociolinguistic Approach", *Forum Linguisticum* I:1, 67-94.

## SOCIAL FACTORS IN THE SOUND CHANGES OF MODERN DANISH

Lars Brink and Jørn Lund, University of Copenhagen, Denmark

In the following we will present some of the major results of our research<sup>1</sup> on the role of social factors in the sound changes of Modern Danish. Our investigations deal with phonetic history based on phonetic sources. The oldest living informants we tested were born in 1875. Edison and others made it possible to investigate an added generation, namely a solid group of informants born from 1840 on. The oldest Danish voice preserved, to our knowledge, must have resounded for the first time in 1813, two years before Waterloo! The recordings comprise at least 10 informants per 5-year period, except for the very first years. Our goal was to register and describe all ascertainable pronunciation changes for those informants raised in Copenhagen and to survey pronunciation and its development outside Copenhagen.

The phonetic history of Copenhagen speech in the period treated reveals an amazing wealth of sound changes. This is largely the result of ca. 60 regular sound laws. We have described the sound changes by consistently referring to the birth date of the informants - since this was our first significant result: When we arranged the material according to the informants' birth dates, numerous sharp boundaries appeared. In fact, often a clear shift between informants born in two subsequent 5-year periods became apparent. If we arranged the material according to recording dates rather than birth dates, no changes could be demonstrated at all, unless the age distribution in the groups was kept painstakingly constant. And even then, the sound changes would only appear as weaker shifts in a sea of variation. - The reason, of course, is that regular sound changes occur between generations (in our opinion innovations

(1) "Dansk Rigsmål I-II. Lydudviklingen siden 1840 med særligt henblik på sociolekterne i København" (Standard Danish I-II. The phonetic development since 1840 with special regard to the sociolects in Copenhagen). 823 pp., Gyldendal 1975; "Udtaleforskelle i Danmark" (Differences in pronunciation in Denmark). 113 pp., Gjel-lerup 1974; "Regionalsprogsstudier" (Studies in regional language), in Danske Studier 1977 (by Jørn Lund); "Om lydlove" (On sound laws), paper presented to the Soc. of Nordic Phil. in Denmark 1977 (by Lars Brink); and an investigation of the linguistic situation in Copenh., 219 pp., 1978 (primarily by Jens Normann Jørgensen). - Here we economize with respect to examples and documentation, and for definitions and methodology we refer to the above works.

come from children, who - under mutual influence - retain while growing up a few of their originally many deviations from the adults' language), and that influence from younger speakers on older ones is modest and never strong enough to allow the older to "catch up" with the younger. We have several recordings of the same speakers made as far as 50 years apart, and they reveal only slight differences. On the whole, it is our experience that most adults who do not change their milieu are only weakly influenced phonetically by the next generation.

It was clear from the start that the material had to be arranged according to sociolects. Certain linguistic features are correlated with high social status (in a certain generation, in a certain area), i.e. the feature becomes more and more common as we progress up the social ladder, e.g. the pronunciation [ʃo<sup>+</sup>'fø+g'] (chauffeur). Others are correlated with low social status, e.g. [ʒæ<sup>+</sup>'fø+g'] (here we employ IPA with unmodified values), and still others are socially neutral. Thus, we could arrive at two, and only two, sociolects in Copenhagen: High and Low Copenhagen, i.e. the languages with (almost) totally high - resp. low - or neutral linguistic features, and, of course, intermediary forms. In the following we will trace the major trends in the development of the two Copenh. sociolects.

Prior to ca. 1750 (birth year), there were practically no social linguistic variations. There were a few differences in the pronunciation of foreign words, and certain folk etymologies must have belonged solely to the lower level. The linguists of the day operate with many divisions: Copenh./provincial, the various dialects, free speech/reading pronunciation, but no one ridicules the common man's deviations from the learned except for certain "distortions" of foreign words. A number of somewhat later authors write, on the contrary, that pronunciations, inflections, etc. which clearly belong to craftsmen and servants in their day were quite common among the learned in previous generations. This agrees with our findings that the numerous certain L-features (L- = low-) in the 1800's can nearly always be traced back to a time when they were common in higher circles. The situation was the same in the country districts. The linguistic interaction in these less specialized societies was probably simply too great to allow social characteristics to arise. A town like Copenhagen

was small (in the 1700's not quite 3 km<sup>2</sup>) and densely populated. The division into better and poorer neighbourhoods belongs to a much later age.

The social uniformity refers to language, not to all aspects of speech,<sup>2</sup> since speech involves many language-independent aspects. Of course, there will always be statistical differences between high and low speech. Due to statistical differences in interests, experiences, intellectual equipment, etc., there will be statistical differences e.g. in conversation topics, sentence length, metaphors, irony, slips of the tongue, syntactic anacolutha, etc. But the fact that L-speakers might discuss boxing more often, or perhaps employ more anacolutha than H-speakers has nothing to do with their language, i.e., the traditionally transmitted set of rules which govern their speech. No linguistic rule recommends or forbids talk of boxing, and it is self-contradictory to maintain that L-language could be viewed as requiring the use of anacolutha, since by this we naturally do not mean deviations in regular speech from regular writing but syntactic deviations in actual speech from regular speech (speech which does not in the least violate the linguistic rules of the speaker). For everyone, their number is great.

The recognition of a social uniformity prior to 1750 must be modified on a few points. The various work spheres have always had a set of terms generally unknown outside the field. This is so obvious that we do not consider it in the notion of a sociolect. If we did there would in every society with a division of labour be just as many sociolects as fields. Other differences in vocabulary involve literary words and learned, foreign words. Finally, in the lower circles there must have been somewhat weaker taboos on swear words and obscene words. None of the two last mentioned situations involve actual sociolect differences: They are even on the highest or lowest social level extremely individually determined. True sociolect features do not reflect individual personality features of the speaker but merely the habits of his surroundings. The decisive argument is that the above mentioned

linguistic features correlate to a higher degree with characteristics such as certain types of knowledge or attitudes toward taboos and only indirectly, and more weakly, with social class. In many cases these non-linguistic or non-sociolectal situations are mixed together with the true sociolect features such that one can obtain the quite distorted impression that high and low language are arranged on a scale, the extremities of which are written academic language and children's speech: Written academic language contains a maximum of regularity, literary terms and urbane expressions - children's speech a minimum.

In the time following, a long series of true sociolect features emerges in pronunciation, inflection, syntax and the core vocabulary. In the beginning, ca. 1750-1800 (birth year), it is a matter of a few features, always such that the lower social levels retain an older form while the higher levels take on a new form or limit themselves to one of two older double-forms. No written source indicates the existence of sound changes of the type: the uneducated have zero where the educated have [h], etc.

Not until ca. 1800 (birth year) do we find this type of differences. The oldest socially staggered sound changes we can ascertain from the phonetic material from 1840 on must have been initiated at the earliest in the generation born in 1800. The first indication in written linguistic sources of such differences we find as late as in the 1880's among authors born after 1850. Apparently no one was aware of these differences for a long time.

From ca. 1900 (birth year) the new changes do not appear nearly as staggered socially, and a series of new sound changes originating in the previous period, both in the L- and in particular in the H-languages, now become reversed. Among the youngest living adults there are still differences left, but on the one hand, these show a tendency toward leveling, and on the other hand, the strength of the social correlations themselves is decreasing. This is all probably due to greater social mobility and integration, and of such individual factors we feel, without being able to prove it, that the most important cause is the fact that from ca. 1900 it became common for the educated to send their children to free, state-supported schools. Radio and television have had no influence, since leveling has most often been in the direction of the L-language, and since in radio and television up to ca. 1970 the H-language was almost always used.

-----  
 (2) in the basic sense "the actual output of speaking" - including all its phonetic and non-phonetic sides.

Of the many regular sound changes appearing after 1800, the majority originate in L-Copenh., e.g. [aɨ > ε] before alveolars and zero, [a-: > ε:], the numerous vowel openings before and after r, [ɣh > ɣ<sup>sh</sup> > ɣs], and the openings of the medial and final spirants [-v > -v̥] [-y > -wɾ/-jɾ] ([wɾ] after back, [jɾ] after front vowels), [-ðɾ > -jɾ] and [-v > -wɾ]. It may seem surprising that the H-sociolect does not lead the way, but this is only at first glance. First of all, it is only natural for new changes to arise in the largest linguistic community. Secondly: for a long time the change is not obvious to speakers, characterized merely by the fact that a growing number of speakers begin to show uncertainty with respect to the new and the old form. The change usually spreads to the H-sociolect one or two generations later where the same situation of uncertainty then appears, and first then do attentive H-speakers become aware of it and, normally, offended, but then it is too late, since the remaining younger H-speakers are pushed from three corners: from the majority of L-speakers, from the minority of H-speakers, and, most importantly, from the "inherent plus-value" of the new forms. It is namely no accident that the new forms could expand from non-existence and thus defy the general imitation tendency (!). - A number of changes, however, originate in the H-language, e.g. the shift of back vowels: [ɔ+ > ɔɨ+], [ɔ+ : > ɔɨ+ :], [ɔɾ > vɨ], [a+ > aɨ], and vowel shortening before vocoids, especially before [ɣ], and some changes show no clear social staggering.

Sociolect and sex. The two sociolects show uneven distribution with respect to sex. Women generally have more H-features than men. Indeed, the group of pure L-Copenh. speakers is made up of more than twice as many men as women; with respect to the pure H-speakers, the difference is not nearly as great. However, we are still dealing with sociolects and not sexolects: the difference between high and low is in every instance investigated greater than the difference between women and men. This can hardly be due to anything but the fact that women for some reason have greater social aspirations than men and are perhaps also more attentive, but these factors need not be the main reason in every concrete instance in which, e.g., the girls in a group of siblings have more H-features than their brothers, since once the sexual difference has come to exist, it can then largely be maintained simply by the fact that girls are more apt to imitate other girls and women.

Orthographic influence. Our investigation shows that regular sound changes, not orthography, are the most important sources of phonetic change. The ca. 60 sound laws which we have ascertained are either neutral towards the orthography or, the majority, directly contrary to it. They have been accompanied by numerous new mergers in vocabulary, e.g. 'ret' = 'rat' (right, steering wheel): [ʰaɨɣ], 'sagn' = 'savn' (legend, lack): [ʰsaɨwɾ'n], 'æder' = 'edder' (eats, venom): [ʰeɾðɾ'vɨ], 'lure' = 'luer' (eavesdrop, flames): [ʰlu+ɾ], and 'løjet' = 'lodde' (lied, solder): [ʰvɨðɾ'ðɾ], and in general this has made spelling more difficult. In addition, many changes in isolated words are contrary to the orthography. We have found many instances of orthographic influence, in proper names, in less common words, and in foreign words in the L-language and also some cases in the core vocabulary; but in ordinary spoken language, in a running text, orthographic influence accounts for a very small portion of all the changes in pronunciation occurred between the 1840 generation and today's youth.

The range of the standard language. In Denmark there exists a non-localizable standard language or rather a complete set of un-localizable linguistic forms, i.e. forms which are not tied to speakers raised in particular areas. By comparison with dialects we have attempted to show that these non-localizable forms, wherever they contrast with the respective dialects, have their historical origin in Copenh. speech. A non-localizable standard language is first completely established when non-localizable forms exist without exception. First with the generation born around 1825 was the last "resistance" to Copenh. forms abandoned, namely when the Copenh. [a+jɾ] and [ɔɾjɾ] diphthongs made their way into the provinces. But just because a complete non-localizable language is available it is not certain that anyone speaks it in its purest form. Actually, and quite naturally, only inhabitants of Copenhagen did so. Not until the generation born in 1880 do we find such informants raised in the provinces, namely H-speakers from the Sjælland market towns. Outside of Sjælland all thoroughly investigated informants possessed certain (but not the same) local features. The local characteristic increases everywhere in the market towns as the social status falls, but even on the lowest social levels the language, today, is much closer to Copenh. speech than to the locality's original dialect, now relegated to the rural

areas and dying out even there. On Sjælland we know of no true dialect speaker born after 1920. - Even the newer developments in Copenh. speech, including the many originating in L-Copenh., spread to the entire country, always somewhat "delayed". Thus, the provinces are characterized both by local forms and by a number of standard-archaisms, archaisms which in the end can become locally bound, namely once the new form has been adopted in (a portion of) the rest of the country. - There exist other centers of development than Copenhagen. In fact, all larger province cities exert an influence on their surroundings. For this reason certain vigorous local forms can now be found in larger provincial areas outside their original dialect area, but primarily these centers function as mediators for the Copenh. features which always influence the larger cities earlier and more forcefully than the smaller.

The Copenh. forms in general, as widespread as most of them are now, are no longer felt by speakers to be Copenh. forms, which they actually are only historically. Their success is thus not due to any capital city prestige. This factor was no doubt important in previous centuries when the non-localizable language was in the process of being established. In the provinces there must exist a certain general high-social atmosphere around them, but not even this factor can be the decisive one, since L-Copenh. forms, as mentioned, also spread energetically. We have attempted to show that logically the language of a capital city can succeed in spreading its forms to the rest of the country purely by contagion, namely according to what we term the Napoleon-principle: the enemy is slain where he is weakest and immediately enrolled in the victor's troops. But, of course, prestige plays a significant role. We feel, however, that the most important emotional attitude toward linguistic forms is one of egocentricity, i.e. preference for habitual forms; spot-checks have shown that most L-speakers actually prefer their own forms to H-forms and provincial-speakers their own provincialisms to standard forms (surely, H-speakers' preference for H-forms is greater, since here habit and prestige-value pull together). But the provincial speakers can naturally get their egocentric resistance shattered under sufficiently massive bombardment with standard forms. - Above all, it must be rejected that the orthography or radio and television have created the standard language. The orthography does not indicate the basic quality of the

sounds; the fact that the standard language possesses the original Copenh. qualities [b, d, g, v, w, t, e, r, o, i:], etc., corresponding to written b-, d-, g-, r-, -d, v-, -v, æ, o, no soul can read from the written picture. Nor does the orthography indicate stød, stress or pitch, and regarding the distribution of sounds the Danish orthography is so archaic that it agrees frequently and significantly with other dialects better than with Copenh. speech (e.g. in the case of -p-, -t-, -k-, hv/v, hj/j, nd/nn, ld/ll, weakly stressed -et; -eg-, -øg-, i, y, u + nasal, geminates). To be sure, the orthography can delay the spread of contradictory Copenh. forms, but it has not been able to stop them. The influence of radio and television has to be modest. The developments mentioned above were well under way (the dialects being long since extinct in all larger market towns, in Aalborg, for instance, already with the generation born around 1800) when radio broadcasting began in 1926, and if this factor was significant today, the degree of local characteristics would not reflect so strongly in part the distance from Copenhagen and in part the degree of urbanization (radio and television consumption does not run parallel to these factors).

Language is a social phenomenon. But language is such a varied activity that it also possesses other aspects. The numerous Copenh. sound changes we have encountered are naturally not the result of normal interaction and imitation, nor of imitation combined with prestige, since the new forms initially have no prestige, they expand from zero-level, and no one realizes initially that he has a new pronunciation, and even less where it comes from. New pronunciations must of necessity possess an inherent plus value in order to progress victoriously. Of course, imitation is one element, but the very dynamics of the development involve the fact that the new pronunciation is initially adopted and adhered to by speakers who far more often hear the old one. The plus value may be connected with errors in perception or production, originating independently with many children (or, theoretically, with one) or with an easier articulation. In other words, when we speak of true sound changes, not just old (standard) forms replacing other old forms in other areas through borrowing, we must not forget that sound change is essentially a non-social phenomenon.

## STRUCTURE ET ASPECTS SOCIAUX DES CHANGEMENTS PROSODIQUES

Ivan Fónagy, C.N.R.S. Paris

1. Nous ne disposons à l'heure actuelle d'aucune description d'un changement prosodique ayant eu lieu dans le passé. Cette absence totale de témoignage a pu être interprétée comme indication, sinon comme une preuve ex silentio de l'immuabilité des formes d'intonation. Elise Richter (1933) attribue la stabilité des formes mélodiques à leur caractère naturel, motivé. Or, rien que la diversité de l'intonation des langues romanes, ou d'autres langues appartenant à la même famille, aurait pu éveiller des doutes au sujet d'un principe de l'immuabilité intonative. Des observations récentes semblent indiquer que les structures mélodiques, vues sous l'angle de la diachronie, sont tout aussi mobiles qu'à l'intérieur d'un système linguistique à un moment donné. Les contradictions qui reflètent, dans les cadres de la synchronie, les changements de phonétique segmentale (Doroszewsky 1935, Fónagy 1956, Labov 1972) caractérisent également la prosodie dans différentes langues.

2. Le long débat, parfois très animé (Gill 1936, Fouché [1936] 1952, 49), qu'a provoqué l'ambiguïté de l'accent en français moderne à partir de la deuxième moitié du siècle dernier (Paris 1862, Passy 1891, Meyer-Lübke 1890) peut être interprété comme une "dramatisation" des contradictions inhérentes au système accentuel. L'accent frappe la dernière et/ou la première syllabe des groupes accentuels en fonction des contraintes segmentales rythmiques, syntaxiques, sémantiques, et surtout, selon le genre du discours, les habitudes professionnelles ou individuelles du locuteur. L'accentuation se présente comme une fonction à variables multiples. Le nombre, l'importance et la valeur de ces variables changent continuellement, et ceci depuis le début du dix-neuvième siècle selon le témoignage des grammairiens (Scoppa 1816, 220).

Toutes conditions égales par ailleurs, il est plus probable que l'accent frappe la première syllabe de l'unité accentuelle si

- a) cette syllabe est fermée;
- b) sa voyelle est susceptible d'être allongée;
- c) le premier mot est un déterminant suivi d'un déterminé - surtout s'il s'agit d'adjectifs numériques;
- d) le premier mot (pronom, adjectif, adverbe interrogatif) figure en tête d'une question partielle;
- e) le mot figure en tête d'une phrase impérative;
- f) ou en tête d'une réponse composée d'un seul mot, etc.

La probabilité de l'accent change, toutefois, radicalement d'un genre du discours à l'autre. L'accent barytonique devient sensiblement plus fréquent dans le discours politique (Duez 1978); la distance moyenne entre syllabes accentuées diminue d'une façon drastique dans la présentation des informations où l'accentuation des mots "enclitiques" est presque érigée en règle. Si la probabilité d'un accent frappant un mot enclitique (préposition, article, conjonction, pronom "atone", verbe auxiliaire) est de 0.03 dans le récit (conte de fées), elle varie entre 0.03 et 0.08 dans la conversation et monte à 0.42 dans les informations télévisées (I. et J. Fónagy 1976).

L'extrême diversité de la distribution de l'accent dans l'énoncé reflète un changement en cours et le masque en même temps. L'accent barytonique apparaît comme l'expression d'un contenu mental (émotion, emphase), comme trait caractéristique d'un genre du discours ou d'un style professionnel, et non pas comme une manifestation d'un changement prosodique. Le rapport de cause à effet est interprété en termes de fin et de moyens. Il est intéressant de voir que même des linguistes distingués n'échappent pas à cette vision finaliste propre aux membres de la communauté linguistique (cf. Fouché [1936] 1952, 51 et s., Fónagy 1979, 180).

Les contradictions socio-phonétiques qui semblent caractériser les changements phonétiques en cours sont l'indice du changement d'accent, mais ne fournissent pas une preuve suffisante du changement. Les premiers témoignages sûrs d'une accentuation barytonique datent du début du dix-neuvième siècle (Fónagy 1979, 168 et ss.). Il faut noter également que les "irrégularités" prosodiques relevées par Richard Strauss chez Debussy (v. ses échanges de lettres avec Romain Rolland), ne se rencontrent jamais chez Lully ou Rameau. Nous avons pu comparer les résultats de tests de perception faits à partir d'enregistrements de 1914-1915 (discours de Poincaré, de Deschanel et de Viviani), d'une part, et d'enregistrements de discours politiques prononcés en 1974, d'autre part. Les mots perçus avec un accent principal ou secondaire sur la première syllabe (sans autres accents), ou ceux ayant un accent principal sur la première et un accent secondaire sur la dernière syllabe, sont nettement plus fréquents dans les discours de 1974, malgré les divergences individuelles à l'intérieur des deux groupes. Cet écart est statistiquement très significatif ( $\chi^2 = 2357.91$ ,  $p < 0.001$ ). On ob-



tient un écart plus faible, mais toujours très significatif ( $\chi^2 = 19.72$ ,  $p < 0.001$ ) en comparant au discours de Poincaré une lecture contemporaine du texte, lecture voulue neutre et qualifiée comme telle par cinq juges. (La valeur moyenne de l'emphase attribuée à la lecture était de 0.85 à partir d'une échelle sémantique de sept degrés, 0 - 7.)

Il n'y a pas de changement évident, par contre, dans la densité, la distribution de l'accent au long de l'axe du temps.

Peu d'indication d'un changement de la fréquence de l'accent en fonction des catégories de mots, sauf: le nombre sensiblement plus élevé des enclitiques accentués dans le corpus de 1974 (cet écart est statistiquement significatif,  $\chi^2 = 330.07$ ,  $p < 0.001$ ).

3. Quant aux changements d'intonation en cours, j'ai pu relever au cours des années cinquante l'apparition récurrente de l'intonation interrogative dans les phrases impératives de sujets hongrois. Cette substitution, cette métaphore mélodique, était particulièrement fréquente dans certains groupes professionnels (contrôleurs de tramway, employés de magasin) d'une part, et chez les jeunes d'autre part (Fónagy 1969). Le transfert était à l'époque motivé, c'est-à-dire limité à des cas où l'intonation montante-descendante des phrases impératives impliquait la présence de certains éléments sémantiques de la modalité interrogative, du moins dans la parole des adultes. Les tests de perception qui ont permis de corroborer ce jugement intuitif montraient en même temps que, pour les plus jeunes, l'intonation interrogative s'opposait comme invitation polie à l'impératif proprement dit, sans avoir nécessairement une implication interrogative. Le transfert intonatif était donc plus fréquent chez les jeunes et en même temps moins précis, moins marqué du point de vue sémantique. Ces divergences dans la fréquence et dans l'interprétation du transfert mélodique selon les générations semblaient indiquer qu'il s'agissait d'une métaphore mélodique figée, donc d'un changement d'intonation en cours.

Les enquêtes récentes faites à partir du même corpus montrent que le changement est en nette progression. Chez les jeunes, l'intonation montante-descendante apparaît comme la forme impérative non marquée qui s'oppose à l'ordre énergique, agressif.

Ces conclusions provisoires sont basées sur des tests faits à partir de variantes synthétisées de la phrase Figyelj ide "Ecoute-moi" (litt.: "Ecoute ici"). Ces enquêtes ne sont pas ter-

minées à l'heure actuelle. J'espère pouvoir en présenter les résultats numériques définitifs au cours du Congrès.

L'intonation terminale montante des phrases assertives dans les dialectes de l'est de la Norvège, signalée par Bertil Malmberg (1966), peut être considérée également comme une métaphore figée. Elle est en effet interprétée comme telle par Kloster Jensen. Selon lui, cette forme d'intonation n'est pas primitive dans ces dialectes: "Elle représente la généralisation d'un type à l'origine stylistique, utilisé pour engager l'interlocuteur, tout comme le nicht wahr? allemand ou le n'est-ce-pas? français." (Je le cite d'après Malmberg 1966, 106).

L'analogie entre métaphore lexicale et métaphore intonative vaut donc également pour la diachronie: dans les deux cas, le transfert aboutit à un changement. Le fait du changement reste inaperçu dans un premier temps à cause de la motivation sémantique du transfert; dans un deuxième temps à la suite de la démotivation qui détache la nouvelle expression de sa base sémantique originale et efface, par là, les traces du changement.

4. Il y a une autre forme de changement mélodique qui ne présente aucune analogie avec d'autres changements linguistiques: c'est l'interférence mélodique, l'union de deux formes d'intonation. Ainsi, une configuration mélodique réunit en français moderne l'intonation interrogative et assertive. Nous avons enregistré, au cours d'un "jeu des portraits" et dans des films policiers, une intonation interrogative-assertive qui peut être considérée comme une forme remaniée de l'intonation déclarative. Elle se distingue de l'intonation interrogative par une chute mélodique finale brusque et de l'intonation assertive par une montée rapide dans la syllabe accentuée, qui est cette fois l'avant-dernière (Fónagy et Bérard 1973). Un phénomène analogue se produit en hongrois où une forme mixte réunit l'intonation interrogative et celle de la protestation indignée (Fónagy 1965). La question incrédule épouse une forme triangulaire en français moderne:

Il e<sup>e</sup>s<sup>t</sup> là<sup>à</sup> ?!

Elle présente, comme la question incrédule américaine (Hadding-Koch et Studdert-Kennedy 1965), allemande, tchèque (Romportl 1973,

153) ou hongroise (Fónagy 1965), une reprise caricaturale de l'affirmation catégorique et interfère en même temps avec l'intonation interrogative (montée dans la syllabe accentuée). Toutes ces formes complexes sont également distribuées dans l'espace socio-culturel; il n'y a donc aucun indice d'un changement en cours. Le changement a dû s'accomplir à une date antérieure. On comprend mieux, dans cette perspective, la coexistence de deux intonations interrogatives en russe (cf. Romportl 1973, 159 et ss)

a) montée et descente dans la syllabe accentuée, vs.

b) montée finale.

La première (a), qui correspond à la courbe mélodique de la question incrédule de l'anglais, du français, etc., figure aujourd'hui comme l'intonation interrogative non marquée (Bryzgunova 1963), et c'est la deuxième (b) qui doit être considérée comme marquée (exprimant l'étonnement ou ayant une valeur d'évocation). Boyanus (1936), qui selon Romportl (1973, 158) était le premier à signaler l'intonation (a), présente la configuration (b) comme modèle de l'intonation interrogative. Il est probable que la configuration (a) était au début, en russe comme dans d'autres langues, une métaphore mélodique remaniée, mais qu'elle s'est généralisée par la suite en se substituant à l'ancienne forme non marquée.

5. Reste à signaler un type de changement mélodique qui suppose une mutation fonctionnelle des formes d'intonation. Une intonation expressive, suggérant une attitude déterminée, devient particulièrement fréquente dans la parole d'un groupe social et finit par se détacher de l'attitude qu'elle est censée exprimer. L'intonation correspondant à une attitude désabusée (une moue, un haussement d'épaules) frappe par sa récurrence tenace dès qu'on écoute les présentatrices de la Radio-Télévision française dans des contextes qui excluent une telle attitude. Nous avons présenté à deux groupes d'étudiants en linguistique (groupe a, groupe b) quatre échantillons de cette forme mélodique, d'abord isolés du contexte (groupe a) puis en contexte (groupe b). Présentés isolément, ces échantillons ont suggéré un air désabusé, voire ironique. Présentés dans le contexte, les mêmes intonations paraissaient à la plupart des sujets comme neutres et ils les attribuaient à une présentatrice de la télévision (I. et J. Fónagy 1976). Ceci revient à dire qu'une forme mélodique expressive est devenue neutre à l'intérieur d'un

groupe professionnel et a acquis par là une valeur évocatrice (Bally 1921, I 203-349). Pierre Léon (1971, 54 et s.), après avoir mis en évidence le caractère irréversible des changements de fonction (expressive → évocatrice, jamais l'inverse), cite comme exemple le style publicitaire composé des traits distinctifs de l'insistance et de la joie. Les formes d'intonation caractéristiques de différents groupes professionnels hongrois (Fónagy et Magdics 1963) se révèlent, dans la plupart des cas, comme des formes d'intonation émotives généralisées, neutralisées.

La généralisation d'une forme peut dépasser tel ou tel groupe professionnel. La montée terminale exprimait en hongrois, et exprime toujours dans certains contextes, certaines attitudes déterminées (surtout l'attitude justificative ou l'expression de l'évidence). Elle est toutefois moins expressive dans la parole féminine où elle prédomine. Elle est aussi en corrélation avec l'âge des sujets et plus fréquente chez les jeunes (figure 1). Au lieu d'exprimer telle ou telle attitude, ou d'évoquer tel ou tel groupe professionnel, elle acquiert un caractère féminin selon des tests de perception (Fónagy et Magdics 1963, 8), moins dans l'opinion des jeunes que dans celle de plus âgés.

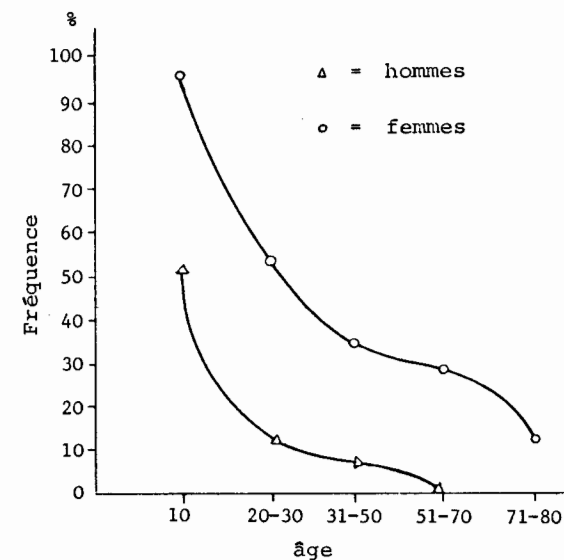


Figure 1

Fréquence de la montée finale dans les questions partielles selon le sexe et l'âge des locuteurs.



6. C'est à force de tels transferts et de telles mutations fonctionnelles que les formes d'intonation de langues apparentées deviennent de plus en plus divergentes d'une langue à l'autre, et que le rapport entre signifiant et signifié devient de plus en plus arbitraire, malgré le lien naturel qui lie l'intonation aux attitudes exprimées.

#### Résumé

Les structures prosodiques, accentuelles et intonatives, sont loin d'être immuables. Leur changement se reflète sur le plan synchronique par la présence de règles contradictoires. L'intonation change par le transfert, l'interférence et la mutation fonctionnelle des formes mélodiques.

#### Références

- Bally, Ch. (1921): Traité de stylistique française I-II, Paris: Klincksieck.
- Boyanus, S.C. (1936): "The main types of Russian intonation", Proc. Phon. 2, 110-113.
- Bryzgunova, E.A. (1963): Praktičeskaja fonetika i intonacija ruskogo jazyka, Moskva: Izd. Mosk. Univ.
- Doroszewsky, W. (1935): "Pour une représentation statistique des isoglosses", Bull. Soc. Ling. Paris 36.
- Duez, D. (1978): Essai sur la prosodie du discours politique, Thèse, Université de Paris.
- Fónagy, I. (1956): "Über den Verlauf des Lautwandels", Acta Ling. Hung. 6, 173-278.
- Fónagy, I. (1965): "Zur Gliederung der Satzmelodie", Proc. Phon. 5, 281-286.
- Fónagy, I. (1969): "Métaphores d'intonation et changements d'intonation", Bull. Soc. Ling. Paris 64, 22-42.
- Fónagy, I. et Bérard, E. (1973): "Questions totales simples et implicatives", Studia Phon. 8, 53-98.
- Fónagy, I. et J. (1976): "Prosodie professionnelle et changements prosodiques", Français Mod. 44, 193-228.
- Fónagy, I. et K. Magdics (1963): "Das Paradoxon der Sprechmelodie", Ural-Altäische Jb. 35, 1-55.
- Fónagy, I. (1979): "L'accent français: accent probabilitaire", Studia Phon. 14, sous presse.
- Fouché, P. (1952): Etat actuel du phonétisme français [1936]; introduction à la phonétique historique du français, Paris: Klincksieck.
- Gill, A. (1936): "Remarques sur l'accent tonique en français contemporain", Français Mod. 4, 311-318.
- Hadding-Koch, K. et M. Studdert-Kennedy (1965): "Intonation contours evaluated by American and Swedish listeners", Proc. Phon. 5, 326-331.
- Labov, W. (1972): Sociolinguistic patterns, Philadelphia: Univ. Press.
- Léon, P. (1971): "Essais de phonostylistique", Studia Phon. 4, Montréal: Didier.
- Malmberg, B. (1966): "Analyse des faits prosodiques - problèmes et méthodes", Cahiers de ling. théor. appl. 3, 99-108.
- Meyer-Lübke, W. (1890): Grammatik der romanischen Sprachen I, Lautlehre, Leipzig: Reisland.
- Paris, G. (1862): Etude sur le rôle de l'accent latin dans la langue française, Paris, Leipzig: Franck.
- Passy, P. (1891): Etudes sur les changements phonétiques, Paris: Didot.
- Richter, E. (1933): "Einheitlichkeit der Hervorhebungsabsicht", Actes Congr. Ling. 2.
- Romportl, M. (1973): "Zum Problem der Fragemelodie", Studies in Phonetics, Prague: Academia (147-164).

## THE SOCIAL ORIGINS OF SOUND CHANGE

William Labov, University of Pennsylvania, Philadelphia, PA, USA

The past century of phonetic research has illuminated our understanding of the production of sounds, the properties of the acoustic signal, and to a certain extent, the perception of speech sounds.<sup>1</sup> Studies of the linguistic organization of these sounds have clarified our understanding of their distribution and diversification, the end results of the process of sound change. But the search for the originating causes of sound change itself remains one of the most recalcitrant problems of phonetic science. Bloomfield's position on this question is still the most judicious:

Although many sound changes shorten linguistic forms, simplify the phonetic system, or in some other way lessen the labor of utterance, yet no student has succeeded in establishing a correlation between sound change and any antecedent phenomenon: the causes of sound change are unknown. (1933:386)

In spite of Bloomfield's warning, linguists have continued to put forward simplistic theories that would attempt to explain sound change by a single formal principle, such as the simplification of rules, maximization of transparency, etc. But at the 2nd Congress of Nordic and General Linguistics, King rejected his own earlier reliance on simplification (1975), and recognized the point made 50 years earlier by Meillet (1921), Saussure (1922) and Bloomfield (1933): that the sporadic nature of sound change rules out the possibility of explanation through any permanent factor in the phonetic processing system. Explanations of the fluctuating course of sound change are not likely to carry much weight unless they take into consideration the parallel fluctuations in the structure of the society in which language is used.

The approach to the explanation of linguistic change outlined by Weinreich, Labov and Herzog (1968) divides the problem into five distinct areas: locating universal constraints, determining the mechanism of change, measuring the effects of structural embedding, estimating social evaluation, and finally, searching for causes of the actuation of sound changes. The quantitative study

-----  
 (1) The results reported in this paper are based on research supported by the National Science Foundation from 1973 to 1978. A more complete report is available in Labov et al., Social Determinants of Sound Change (1978).

of sound change in progress by Labov, Yaeger and Steiner (1972) located three universal constraints on vowel shifting, a line of investigation originally foreseen by Sweet (1888), and expanded the view of functional embedding in phonological space outlined by Martinet (1955). Our current studies of sound change in progress in Philadelphia have developed further techniques for the measurement and analysis of vowel shifts, with the end in view of attacking the actuation problem itself. We have approached the question of why sound changes take place at a particular time by searching for the social location of the innovators: asking which speakers are in fact responsible for the continued innovation of sound changes, and how their influence spreads to affect the entire speech community.

It is often assumed that sound change is no longer active in modern urban societies, and that local dialects are converging under the effect of mass media that disseminate the standard language. The results of sociolinguistic studies carried out since 1961 show that this is not the case: on the contrary, new sound changes are emerging and old ones proceeding to completion at a rapid rate in all of the speech communities that have been studied intensively. Evidence for sound changes in progress has been found in New York, Detroit, Buffalo, Chicago (Labov, Yaeger and Steiner 1972), Norwich (Trudgill 1972), Panama City (Cedergren 1973), Buenos Aires (Wolf and Jiménez 1978) and Paris (Lennig 1978). This evidence is provided by distributions across age levels (change in apparent time), and by comparison with earlier phonetic reports (change in real time), following the model of Gauchat 1904 and Hermann 1930.

Whenever these changes in progress have been correlated with distribution across social classes, a pattern has appeared that is completely at variance with earlier theories about the causes of sound change. If one looks to the principle of least effort as an explanation, or to discontinuities of communication within urban societies with accompanying isolation from the prestige models, then it would follow that sound change arises in the lowest social classes. Arguments for the naturalness of vernaculars and the marked character of prestige dialects would also look to the lowest social class as the originating site of sound change (Kroch 1978). If the theorist focuses on the laws of imitation (Tarde

1873) and the borrowing of prestige forms from centers of higher prestige, then it would follow that new sound changes will be the most advanced in the highest social classes. Neither of these cases has appeared in the internal changes studied in urban societies. It is true that older sound changes, like stable sociolinguistic variables, are often aligned with the socioeconomic hierarchy, so that the lowest social class uses the stigmatized variant most often, and the highest social class least often. But new sound changes in progress are associated with a curvilinear pattern of social distribution, where the innovating groups are located centrally in that hierarchy: the upper working class, for example, or the lower middle class.

Thus in New York City, lower middle class groups were the most advanced in the raising of long open o in lost, law, etc (Labov 1966, 1972). The same pattern was found in the backing of (ay) and the fronting of (aw) in that city. In Norwich, Trudgill found that the backing of short e before /l/ in belt, help, etc., showed a rapid development among younger speakers, and was most advanced in the upper working class (1972). In Panama City, Cedergren found that one of five sociolinguistic variables studied showed an age distribution characteristic of sound change in progress: the lenition of (ch) in cerca, muchacha, etc. This sound change showed a strong peak in the centrally located Classes II and III that Cedergren had established in Panama City (1973).

Our project on linguistic change and variation selected Philadelphia as a site for the further study of this problem, since it appeared that almost all of the Philadelphia vowels were in motion, and all of the basic patterns of chain shifting found in English and French dialects could also be located in Philadelphia. The main data base for the Philadelphia investigation is a series of long-term neighborhood studies in working class, middle class and upper class areas, involving repeated interviews and participant observation of the speech community. To this is added a geographically random survey of telephone users employing short, relatively formal interviews. The convergence of the findings from these two data bases, which show opposing strengths and sources of error, provides strong support for the general findings, though only data from the neighborhood studies will be presented here.

The measurement of vowel nuclei was carried out by a frequen-

cy analysis using a real-time spectrum analyzer (SD 301C), followed by linear predictive coding of the frequency domain (Markel and Gray 1976, Makhoul 1975) to derive more exact estimates of the central tendencies of F1, F2, F3 and F $\emptyset$ . Complete vowel analyses of spontaneous speech were carried out for 97 subjects in the neighborhood studies and 60 subjects in the telephone survey, with 150-200 vowels measured for each subject. The mean values for each subject were then submitted to three normalization programs: a log mean model developed by Nearey (1977), the vocal tract scaling of Nordström and Lindblom (1975), and a three parameter method developed by Sankoff, Shorrock and McKay (1974).

Stepwise regression was carried out on the unnormalized and normalized series, deriving equations that predicted mean F1 and F2 positions from age, sex, social class, social mobility, ethnicity, neighborhood, communication patterns and the influence of other languages. The regression program enters into the equation the independent variable that has the highest partial correlation with the mean formant values, and with each successive term re-examines all previous terms as if they were the last to be added to the equation: if their effect falls below a given level of significance, they are removed (Draper and Smith 1966, Efroymson 1960). Thus the relative order in which variables are presented to the program is immaterial.

We then searched for the method of normalization that showed the maximum clustering to eliminate the effects of differences in vocal tract length, and the minimum tendency to eliminate variation known to be present in the data by independent means. Uniform scaling based on the geometric or log mean (Nearey 1977) was selected by these criteria and will be used as the basis for the discussions to follow.

Figure 1 shows the mean positions of the Philadelphia vowels of 93 speakers in the neighborhood series. It also shows vectors representing the significant age coefficients of the regression equations. The age coefficients are multiplied by the chronological age of the subject, e.g.

$$F2(aw) = 2086 - 5.39 \cdot \text{Age}^{[t=6.0]} \dots$$

where the numbers may be read as F2 values in Hz. Thus the first and most significant coefficient shown above predicts that the difference in mean F2 positions for two speakers 50 and 25 years old

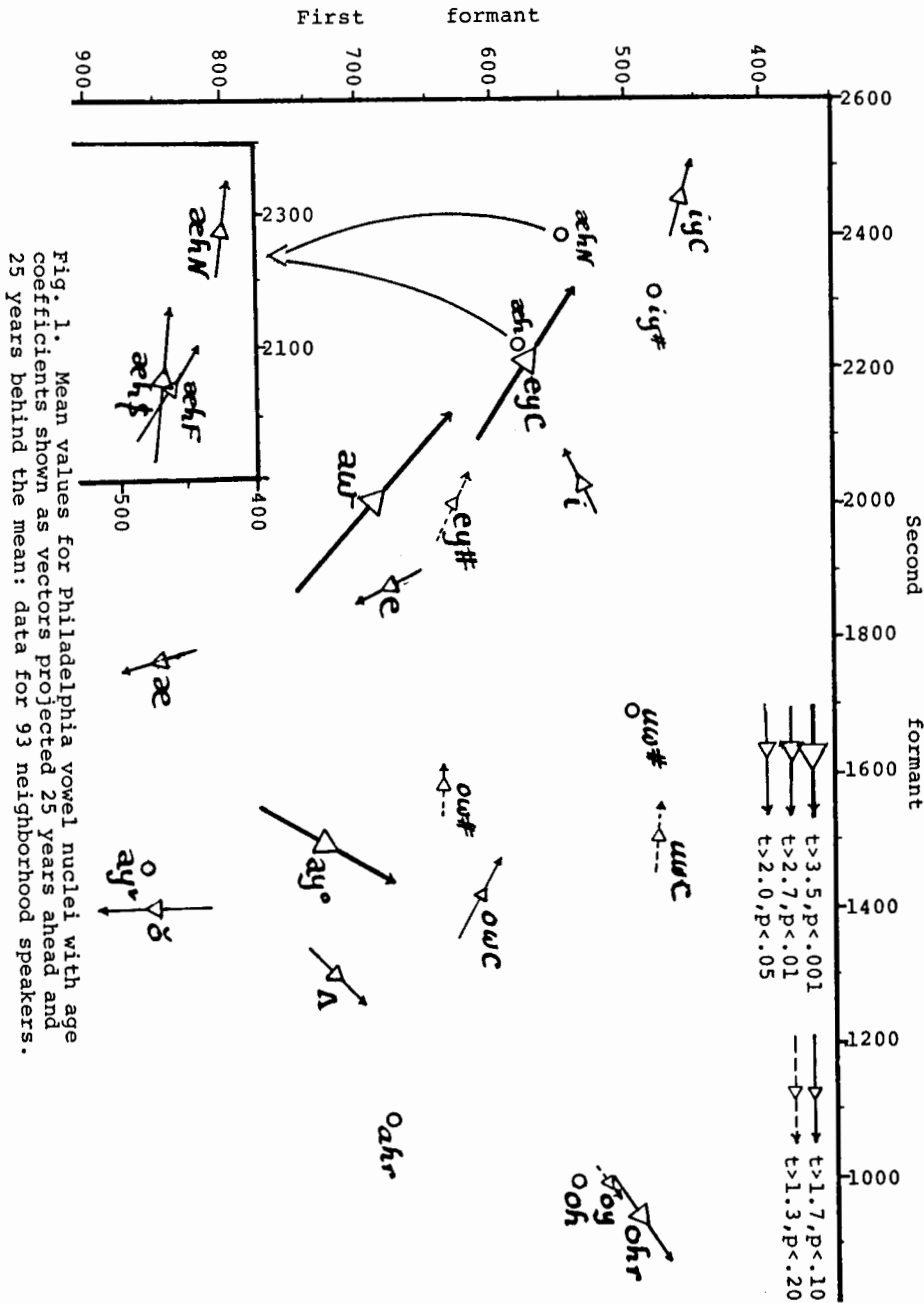


Fig. 1. Mean values for Philadelphia vowel nuclei with age coefficients shown as vectors projected 25 years ahead and 25 years behind the mean: data for 93 neighborhood speakers.

will be (25)(5.39) Hz: that is, the younger speakers will have a mean F2 135 Hz greater than the older. The vectors on Figure 1 represent the result of projecting the sound change 25 years ahead of the mean value and 25 years behind it. The significance of the effect is shown by the size of the triangles and the heaviness of the vector lines.

These age vectors fit in with evidence derived from earlier records and synchronic characteristics of the current data that allow us to set up five strata of sound change in Philadelphia:

- a) recently completed changes: e.g., the raising of /ahr/ in car, part, etc.
- b) changes nearing completion: e.g., the raising and fronting of (aeh) in man, hand, etc.
- c) middle range changes: the fronting of (uw) and (ow) in too, moved, go and code (but not before liquids).
- d) new and vigorous changes, not reported in earlier records: the raising and fronting of (aw) in house, down, etc., from [æ<sup>u</sup>] to [e<sup>o</sup>]; the raising and backing of (ay<sup>o</sup>) before voiceless consonants in fight, like, etc., from [a<sup>l</sup>] to [ɛ<sup>l</sup>]; the raising of (eyC) in the checked syllables of made, lake, etc., from [e<sup>l</sup>] to [e<sup>ɪ</sup>].
- e) incipient changes, e.g., the lowering of the short vowels /i/, /e/ and /æ/.

Conclusions from the earlier studies would lead us to associate a curvilinear social pattern with (d) the new and vigorous changes represented by the long, heavy vectors in Figure 1. Further terms in the regression equation show that this is the case. Extending the equation for F2 of (aw) to the three next most significant coefficients, we have [SEC = 'socio-economic class']:

$$F2(aw) = 2086 - 5.39 \cdot \text{Age} + 126 \cdot \text{Female}^{[t=3.5]} + 261 \cdot \text{SEC } 9^{[t=3.1]} - 253 \cdot \text{SEC } 13-15^{[t=2.5]}$$

In this socio-economic class scale, a 16-point index based on education, occupation and residence value, SEC 9 is generally considered the highest section of the working class, and 13-15 the upper middle class. Fig. 2 shows the coefficients for F1 and F2 projected as a single index on the front diagonal for all SEC, forming a smooth curvilinear pattern around SEC 9. Non-significant points are consistent with the main effects shown above.

Figure 3 shows the class distribution for the projection of checked (eyC). This is a broader curvilinear pattern with a significant peak in the middle working class group SEC 7, and two

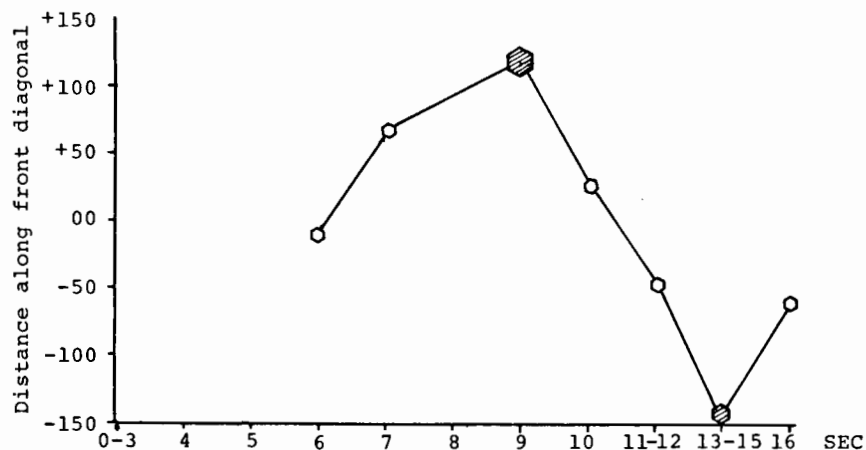


Fig 2. Projection along front diagonal of regression coefficients for F1 and F2 of (aw) for all socio-economic classes compared to SEC 0-3: data from 93 speakers in Philadelphia neighborhood study.

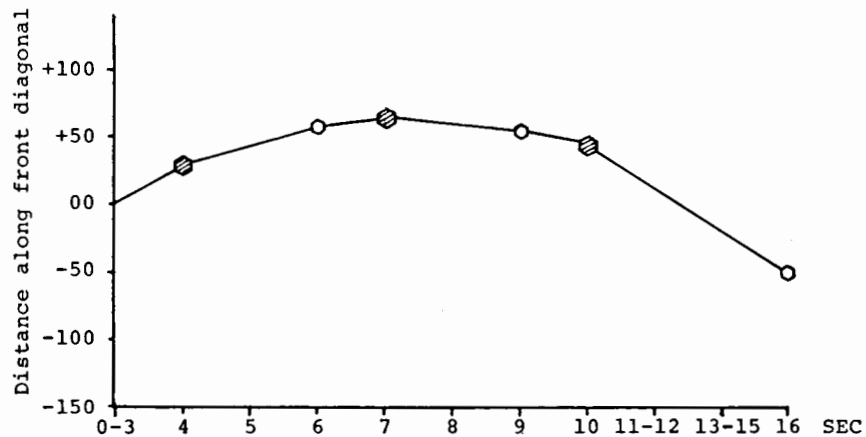


Fig 3. Projection along front diagonal of regression coefficients for F1 and F2 of (eyC) for all socio-economic classes compared to SEC 0-3: data from 93 speakers in Philadelphia neighborhood study.

- ▨  $p < .01$
- ◐  $p < .05$
- $p < .10$

other points significantly higher than the reference level of SEC 0-3, located symmetrically above and below SEC 7. Again, the less significant points form a smooth curvilinear pattern.

The third new and vigorous change, the raising of (ay<sup>0</sup>), shows no significant class distribution. It is worth noting that this is also the only change where men are in the lead: as in most previous studies of vowel change Philadelphia women are about one generation ahead of men in the early stages of change, except in the case of (ay<sup>0</sup>). Whatever the explanation for this connection between sex and SEC patterns, the Philadelphia results agree with impressionistic studies in showing no cases where the lowest or highest social classes appear as innovators in systematic change.

Given the powerful tendency for systematic sound changes to arise in interior social groups, we must ask how this observation bears on the causes and motivations of sound change. Instead of pursuing speculations on the psychological traits of these upper working class innovators, it will be more fruitful to probe more deeply into their social roles and relations to others in the community. The further investigation of the problem carried out by our research group is based on the evidence of communication networks which cannot be presented in this brief report. In general, it can be said that the speakers who are most advanced in these sound changes are those with the highest status in their local community: more specifically, they are persons with the largest number of local contacts within the neighborhood, yet at the same time with the highest proportion of their acquaintances outside the neighborhood. A portrait is beginning to emerge of the individuals with the highest local prestige who are responsive to the broader, almost metropolitan prestige that has become associated with the sound changes in question. It is plain that we are dealing with the emblematic function of phonetic differentiation: the identification of a particular way of speaking with the norms of a particular local community.

Through the further study of the role of new ethnic groups entering the community, and the communication patterns that connect local neighborhoods, we hope to delineate more closely the social pressures that are responsible for the dissemination and further advance of sound change, and thus isolate the driving force behind the continued diversification of linguistic forms.

## References

- Bloomfield, L. (1933): Language, New York: Henry Holt.
- Cedergren, H.J. (1973): The interplay of social and linguistic factors in Panama. Unpublished Cornell U. dissertation.
- Draper, N.R. and H. Smith (1966): Applied regression analysis, New York: Wiley.
- Efroymson, M.A. (1966): "Multiple regression analysis", in Mathematical methods for digital computers, A. Ralston and H. S. Wilf (eds.), 191-203, New York: Wiley.
- Gauchat, Louis (1905): "L'unité phonétique dans le patois d'une commune", In Festschrift Heinreich Morf, 175-232, Halle: Max Niemeyer.
- Hermann, E. (1929): "Lautveränderungen in der Individualsprache einer Mundart", Nachrichten der Gesellsch. der Wissenschaften zu Göttingen, Phil.-his. Kl., 11, 195-214.
- King, R. (1975): "Integrating linguistic change", The Nordic Languages and Modern Linguistics, K.-H. Dahlstedt (ed.), 47-69, Stockholm: Almqvist & Wiksell.
- Kroch, A. (1978): "Toward a theory of social dialect variation", Language in society 7, 17-36.
- Labov, W. (1966): The social stratification of English in New York City, Washington: Center for Applied Linguistics.
- Labov, W. (1972): Sociolinguistic patterns, Philadelphia: University of Pennsylvania Press.
- Labov, W., M. Yaeger and R. Steiner (1972): A quantitative study of sound change in progress, Philadelphia: U.S. Regional Survey.
- Labov, W., A. Bower, D. Hindle, E. Dayton, M. Lennig and D. Schiffrin (1978): Social determinants of sound change, Philadelphia: U.S. Regional Survey.
- Lennig, M. (1978): Acoustic measurement of linguistic change: the modern Paris vowel system. Unpublished University of Pennsylvania dissertation.
- Makhoul, J. (1975): "Spectral linear prediction: properties and applications", IEEE Transactions on Acoustics, Speech and Signal Processing Vol. ASSP-23, No. 3.
- Markel, J. and A. H. Gray, Jr. (1976): Linear prediction of speech, Cambridge, MA: Bolt, Beranek and Newman.
- Martinet, A. (1955): Economie des changements phonétiques, Berne: Francke.
- Meillet, A. (1921): Linguistique historique et linguistique générale, Paris: La société linguistique de Paris.
- Nearey, T. (1977): Phonetic feature systems for vowels. Unpublished University of Connecticut dissertation.
- Nordström, P.-E. and B. Lindblom (1975): "A normalization procedure for vowel formant data", Paper 212 at the 8th Int. Cong. of Phonetic Sciences, Leeds.
- Sankoff, D., R. W. Shorrock and W. McKay (1974): "Normalization of formant space through the least squares affine transformation", Unpublished program and documentation.
- Saussure, F. de (1922): Cours de linguistique générale, 2nd ed., Paris.
- Sweet, H. (1888): A history of English sounds. Oxford: Clarendon Press.
- Tarde, G. (1873): Les lois d'imitation.
- Trudgill, P. J. (1972): The social differentiation of English in Norwich, Cambridge: University of Cambridge Press.
- Weinreich, U., W. Labov and M. Herzog (1968): "Empirical foundations for a theory of language change", in Directions for historical linguistics, W. Lehmann and Y. Malkiel (eds.) 97-195, Austin, Tex.: University of Texas Press.
- Wolf, C. and E. Jiménez (1978): "A sound change in progress: the devoicing of Buenos Aires /ʒ/ into /ʒ̥/", unpublished paper.

## SOCIAL STRUCTURE AND PHONEMIC MODIFICATION

Bertil Malmberg, Dep. of General Linguistics, Östervångsvägen 42,  
223 65 Lund, Sweden

When I made my first efforts to apply principles of Prague phonology in an analysis of the French and Italian vocalic systems (Acta Linguistica II, 1940-41, 232-246; III, 1942-43, 34-56), I soon arrived at the conclusion that this was not possible if I assumed that two units of expression (in my case two "vowels"), in a given position, and throughout the vocabulary, were either variants (allophones) or invariants (phonemes). It turned out that certain units were definitely phonological in some words, mere variants (free allophones) in others. The fact that /e/ - /ɛ/ in final position is definitely distinctive in pairs like dé-dais, fée-fait does not exclude their use as free variants in e.g. quai, gai, (je) sais. As far as the two a:s (/a/ and /ɑ/) are concerned, they have their full phonological value only in relatively few words (lâ-las). Even Parisians (only the language of the capital is referred to here) who agree on the existence of the opposition often do not agree on the distribution of the units in the vocabulary. I had also mentioned in my early study the critical oppositions /ø/ - /œ/ and, though better maintained, /o/ - /ɔ/ (both in closed syllable). I also mentioned the quantitative opposition, still retained by a few speakers, between /ɛ/ (mettre) and /ɛ:/ (maître). The cases of merger were far too frequent to be dismissed as mere phonemic word variants (Jones). I had drawn from these findings the conclusion that it was reasonable to look upon the French vocalism as containing two phonological systems, one richer and another poorer, or in my terms, a maximum system and a minimum system, one of them applied by certain speakers and in certain types of words, the other applied by others and in other words. I never concluded that some speakers used just one, others the other of those two systems in their entirety. I still do not know if there are native Parisians who make full use of the maximum system in any possible position and other native speakers who content themselves with the minimum one throughout the vocabulary. What is certain is, however, that the latter case seems to be normal in the pronunciation of numerous immigrants from the provinces and particularly in southerners and French immigrants from North Africa.

There seems to be no doubt that the choice between a more complex and a more reduced system is determined by non-linguistic (social, cultural, and in the case of immigrants from other French-speaking areas, regional) factors. The maximum system represents the complete set of oppositions permissible according to the paradigm and all the syntagmatic distinctions admitted by the distributional laws - the minimum system the smallest number of distinctive units without which the message does not function and the identification of the meaningful units ceases. (When putting it that way I do not take into consideration factors such as redundancy and context.) In other words, the difference between the two is one between what a speaker can and what he must do. The same interpretation seemed to me to be useful in the analysis of other complicated systems, i.e. the word accent problem in Scandinavia and (as demonstrated in the article quoted in Acta Linguistica III), the oppositions /e/ - /ɛ/ and /o/ - /ɔ/ in Italian, where in both cases it is a question of interference between dialects (or regional variants of the standard), whereas in French the situation can at least partly be interpreted as one between diachronically different systems (though both present at the same time and transformed into social or individual phenomena). Consequently, a state of language (introduced here as a translation of état de langue used in my French text,<sup>1</sup> a concept which goes back at least as far as Saussure's "Cours") may contain different strata, the most simplified of them pointing in the direction the evolution will take if no intervening factors prevent it. It is from this point of view that such an idea may be useful for a correct interpretation of diachronic, or evolutionary phonology. A language thus becomes a harmonious achronic system, or rather complex of systems, whereas a state of language is a linguistic situation described as valid for a chosen period of time or/and for a chosen spatial region or social stratum (all arbitrarily chosen).

The minimal system of French vowels represents a reduction in relation to the fuller one; in purely synchronic terms a system of inferior complexity. Diachronically it represents a loss of certain oppositions retained in the richer one. In all the cases under

(1) See my article in *Mélanges Straka I*, 1970, reprinted in Malmberg, "Linguistique générale et romane", Mouton, Paris 1973, 155-159.



discussion, the distinctions are phonetically subtle. This means that the oppositions based on the slightest differences of articulation and perception have been eliminated or, in most of our examples, reduced in their usage to a small number of words, forms, and contexts. This is typical of what happens in languages in reduction or destruction (in evolutionary phonology, in aphasia, etc.), and in reversed order in languages in construction (in the child, in the language learner, etc.). This is a consequence of Jakobson's law, implying that the complex system supposes the less complex ones, the subtle differences the rougher ones. We know that this law is valid in language learning and in language loss. It must necessarily be taken into consideration also in a study of linguistic change (phonological or other). We also know that the complete elimination of a language - under the pressure of another or owing to lack of motivation for its conservation - takes place according to the same hierarchic order. A situation such as the one reflected in the actual French vocalism is typical of a stage which precedes a generalized simplification. This does not mean that the simplification will necessarily take place. The choice of the speakers may be directed towards a retention of status quo, or even lead to a reestablishment of the more complex system (an example seems to be the opposition /e:/ - /ɛ:/ in the Swedish pronunciation of Stockholm).

My thesis is consequently that any state of language contains levels of different complexity from the maximum system maintained by strong linguistic norms, through degrees of increasing simplification down to the minimum system, and even beyond these to defect forms of language in the child, in aphasia, or in other disorders such as deafness, and in such foreigners and bilinguals as belong only partly to the socio-linguistic group in question. Any language system and, more generally, any semiotic system, is maintained thanks to a tradition respected by the members of society. Its basis is the prestige of norms regulating people's behaviour. The structural reduction of a system and its final elimination is the inverted function of the strength of the norms which guarantee its validity. Consequently, the existence of levels of varying structural complexity is due to the incapacity of the norm to maintain the complete system down to the lowest strata of society, in the more distant parts of the linguistic community, and under un-

favourable external conditions. Those are only aspects of the same phenomenon. In earlier studies and particularly with reference to Romance and Hispanic phonological evolution,<sup>2</sup> I have proposed to talk about simplification in the periphery. It follows from what has been said so far here that the concept of periphery is used with reference to two dimensions: spatial and social. The simplified or defect linguistic usage in the lowest social and cultural strata is peripheral in the same sense as the form of language in distant regions, far from normative centres. The concept of distance is consequently taken as meaning horizontal as well as vertical remoteness. With a slightly deviating use of the term it may even be extended to cover a weak (individual) mastery of the functional system.

If we look upon a state of language as a unity of systems of varying complexity, it will be necessary to introduce as a further variable the concept of choice. A Frenchman of today may choose one type of structure or another. His choice will be determined by his preference for one or another of existing norms (any linguistic usage being, of course, governed by some norm). He may make his choice unconsciously and in accordance with his social (cultural) background, or in a conscious intention to manifest his position as belonging to the upper ten, or as loyal to the social group where he comes from or to which - for personal or ideological reasons - he wants to belong. In such cases, his choice of pronunciation may function in exactly the same way as his choice of clothes or his social behaviour in general.

When, in my plenary report to the International Congress of Linguists in Bucharest (in 1967), I formulated the consciously and intentionally provocative thesis that language does not change and that what we call linguistic change is the speaker's choice of another language (taken here as a stable system of functions, and independent of any time factor), I thereby wanted to stress the importance of the choice factor in the evolution. I found it fruitful to see language as consisting of strata or levels, the choice between which is determined by social evaluations, even by changing

-----  
 (2) Summarized in *Orbis* XI, 1, 1962, 131-178 (reprinted in "Phonétique générale et romane", Mouton, Paris 1971, 301-342), and, as far as Spanish is concerned, in "La América hispano-hablante", Istmo, Madrid 1970.



modes. We have seen that modifications by choice may in principle take place in two directions: downwards, towards a simpler structure, and upwards, i.e. replacing a simpler structure by a more complex one. The danger of homonyms, often quoted as an important positive factor, and the absence of them as a negative one, has probably been exaggerated.

Interference (substratum, superstratum, adstratum) has often been quoted as an underlying factor in sound change. It supposes bilingualism. Bilingual areas and societies are given as examples of conditions under which the linguistic norm may be weakened and where system reductions a priori seem probable (well-known examples are the loss of voiced stops in the French spoken in Alsace and the loss of the phonemic word accent in the Swedish of Finland). Now an important question arises: are such peripheral simplifications to be explained through direct influence from the language which ignores the distinction, or are they simply due to a general weakening of the norms in a peripheral area? We know that voiced stops are relatively rare and that they come late in the child's linguistic development. We also know that phonological word tones belong to the subtle phonological distinctions, late in Swedish children and absent in cases of individual linguistic weakness. This question can hardly be answered. The effects are the same. When the change is just a phonological reduction, the interference theory is superfluous. Only when the new system contains new structural features and/or structural relations do we have any real reason to consider an interference theory. The introduction into Northern Gallo-romance of the phoneme /h/ as a consequence of the Frank colonisation (retained till today in some dialects, Normandy) is inexplicable without the foreign influence (and understandable in consideration of the socio-linguistic situation in the bilingual Frank kingdom).

It seems, on the other hand, quite normal if in a language in close contact with a quite different neighbouring one whose influence on the former is understandable (socially, culturally, politically, or simply through a quantitative dominance), we meet phenomena of phonetic realization of the phonological system which have to be explained through interference between different speaking habits. The examples are numerous (the lack of aspiration of Swedish-Finnish /p, t, k/; the pronunciation of the Spanish /j/-

phoneme as /d̥j/ in Paraguay; intonation and stress phenomena). These features do not belong to the phonological system strictly speaking (though they may play a part in communication on other levels than the strictly cognitive one). And they may come to play a role at later stages in the phonological evolution (an example later).

In my critical studies on Romance diachronic phonology, I have been very restrictive as far as interference theories are concerned. I have tried to prefer internal evolution and peripheral simplification as explanations, the latter socially determined.

The expansion of Castilian in medieval Spain which became a consequence of the reconquest ("reconquista") from the Arabs, as well as its continuation (from 1492) in America, implied numerous instances of structural simplification of the phonological system (loss of the medieval opposition between voiceless and voiced fricatives, voiced and fricative stops). This evolution was parallel with the social changes brought about by the political events. The medieval /ts/ was replaced by the interdental /θ/ in the centre but confused with /s/ in the South and in America ("seseo"). A widespread dialectal confusion of liquids is found in (regionally and/or socially) peripheral strata all over the Spanish speaking world. It results in a substitution of one for the other (mostly a generalization of l), or in a phonetically intermediate type. A map published by Alonso-Lida (Rev. Fil. Hisp. VII, 1945, 320) of the extension of the merger in Spain shows its marginal character. Other phenomena of simplification show a corresponding spatial and social extension on both continents. The Spanish of America reflects the differences of political, social, and spiritual structure in the colonial period. The replacement of implosive -s in Spanish through an undifferentiated h-like fricative has the same extension as other "vulgarisms", in Spain and in America. Though it is a mere manifestation of the s-phoneme, it may have secondary phonemic consequences (lengthening of vowels, change of vowel quality) and ought to be mentioned for this reason. A parallel evolution took place in medieval French and is still reflected in oppositions like Fr. patte - pâte.

Linguistic evolution would not be conceivable without the hierarchic differentiation of a state of language, without vari-

ability in the strength of norms, and without a choice (free within limits) made by the members of the different social strata. These are the essential factors in the socio-linguistic evolution.

In conclusion: diachrony interpreted as a substitution of one system for another (in any of the dimensions of language) through a socially determined choice between possibilities of varying complexity was the principle I wanted to submit for consideration to the Bucharest Congress of 1967. I did it by saying: language does not change; man changes languages.

## THE REALITY OF SOUND CHANGE: A SOCIOLINGUISTIC INTERPRETATION

Fred C.C. Peng, International Christian University, 10-2, 3 Chome, Mitaka, Tokyo, 181, Japan

This paper attempts to summarize the latest findings of my research on sound change. It also contains criticisms of and comments on previous studies along this line. In the main, a new theory is proposed, suggesting that the process of sound change can be observed within one generation. Given this theory, four questions are asked, which become the focus of my argument in the course of discussion.

The Problem

Sound change has been an intriguing subject in general linguistics for almost two centuries. I wish to emphasize, however, that language as a code does not change by itself; people who employ the code change it. It is from this point of view that I shall address myself to the reality of sound change.

To begin with, let me identify the problem. Linguists have in the past been led to believe that it will take generations to produce certain changes and that the length of time that is needed to show such changes is too long, or to put it the other way around, that the ongoing progress of such changes is too subtle and slow to allow any direct observation.

Labov has recently challenged this traditional belief by advocating that change can indeed be directly observed. However, Labov's observation of sound change in Martha's Vineyard, involving a claim that sound change may be captured while in progress, takes in three generations (Peng 1976, 70), thereby yielding to the "myth" in the literature that changes occur across the boundaries of two or more generations.

This myth was repeated once more by Johnson recently (1976) who claims that "The time span considered can be across several centuries or as few as two demographic generations" (1976, 165). He thus concludes that "Specifying the terms 'fast' and 'slow', we have given some support to the claim that change begins slowly and accelerates in succeeding generations, and we have given evidence that change advances more rapidly in urban than in rural communities" (1976, 171).

In view of this (unfortunate) development, several questions need to be raised here, so as to eradicate once and for all the

myth that seems to persist in the literature. For the sake of convenience, these questions are asked below in the order in which I shall discuss them in this paper:

- Q1. Is it linguistically plausible to construct a theory of sound change that is based on the assumption that sound change takes place across the boundaries of two or more generations?
- Q2. Is it true that change begins slowly and accelerates in succeeding generations?
- Q3. Is it theoretically sound to generalize from one type of changes in one language to the same type of changes in other languages?
- Q4. Can linguists, historical linguists in particular, do themselves justice by ignoring nonlinguistic changes when they deal with linguistic change?

Previous Study on Sound Change within One Generation

Let me quickly review what I said in Peng (1976) concerning sound change within one generation. First I took Nomoto's 1950 study and 1971 study and came up with the result that each individual seems to continue developing his or her speech beyond 13 years of age, at an ever decreasing rate, until the age of 35 or thereabout. I added that "Such is the case in most of the phonetic parameters" (1976, 82). I then proceeded to ask a question: If changes can be directly observed to take place within one generation, what are the mechanisms of sound change that may be discerned from the study? Five mechanisms were then singled out: Age factor, Educational background, Phonetic parameter (i.e., the choice of speech sound), Oscillation (for what Weinreich called retrograde), and Life expectancy. Each mechanism was elaborated on the basis of supporting data (1976, 83-90).

Second, I took Jespersen's metaphor and compared it with my alternative schematic representation of language change, which may be recapitulated as follows (1976, 91):

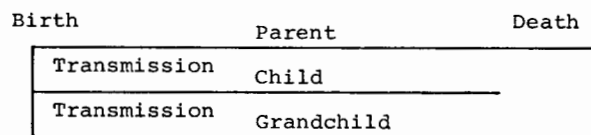


Figure 1. Illustration of Jespersen's Metaphor

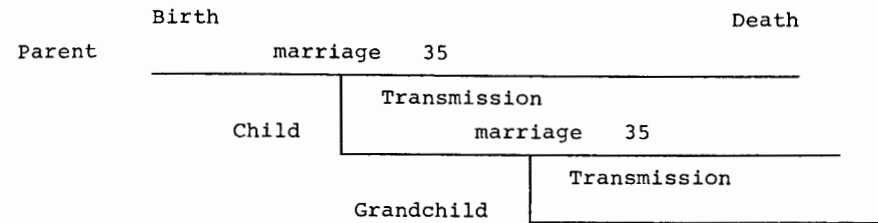


Figure 2  
Alternative Schematic Representation  
of Language Change

This alternative theory suggests that the child can learn his language perfectly; that in spite of his perfect learning, the language in question still changes because the model the child learns his language from had changed considerably before the child was born; and that the child's model had, in turn, learned from quite a different model, just as the child will serve as quite a different model to his own child. In this way, I concluded that a sound change, be it abrupt or not phonetically, can only be gradual in terms of behavior within each individual, with smooth (i.e., perfect) transmission from generation to generation (1976, 92).

To illustrate this point, let me make a distinction between changes in language behavior and changes in linguistic code. This distinction is important because the accumulation of changes in language behavior results in changes in linguistic code, and changes in language behavior are directly observable; on the other hand, changes in linguistic code may or may not be so observed if one's aim is to determine the end points, rather than the ongoing processes, of the period of operation of a sound change. An exemplification of this distinction is in order here.

In his criticism of the 'gradual view', Wang cites an interesting case as follows: "so, for a word like acclimate in which the pronunciation changes from [əkliájmit], the only pronunciation found in some older dictionaries, to [əkliúmejt], where all three vowels are different (in addition to the difference in accent pattern), it is surely unrealistic to suppose that there was a gradual and proportionate shift along all four phonetic dimensions" (1969, 14).

Note that while the change from the first to the second pronunciation may be abrupt along all four phonetic dimensions or even one phonetic dimension, it is notwithstanding a change in the

system of the linguistic code. Thus, the abruptness is immaterial here, because any native speaker of English can switch instantly from one pronunciation to the other with little difficulty.

By contrast, however, the change in language behavior from the first to the second pronunciation must be gradual. This aspect of gradualness can be directly observed and measured as part and parcel of language behavior, among various groups of people with varying social backgrounds.

From the above review it must now follow that if changes in language behavior can be systematically described, there is no need to wait for the result (i.e., the end point) to show up in the code itself. We must come to grips with the ongoing process of changes in language behavior that underlie the net result of changes (i.e. end points) in the linguistic code.

#### Discussion

With the conception of sound change presented above in mind, let me now return to the questions originally asked. First, I must mention that it is rather unfortunate that Johnson repeats the traditional view that sound change must take place across generation boundaries.

Empirical evidence is presented in Peng (1976 and n.d.) that sound change takes place not only within each individual but at an ever decreasing rate, that is, taken cross-sectionally, a person may change his linguistic system within his life span but gradually reduces his rate of change until the age of 35, even though changes may continue to take place after the age of 35 (but at a much reduced rate). In light of this finding, it is hard to believe that sound change must take place across generation boundaries.

Second, given the above finding that sound change takes place within each individual at an ever decreasing rate, I must now ask whether it is true that change begins slowly and accelerates in succeeding generations. Although the data presented by Johnson may seem suggestive of this tendency, a closer look at his data indicates otherwise (especially when they are compared with ours), simply because ours can account for changes within one generation, whereas Johnson's (which include several sources) contain materials from at least three generations, each having a different age bracket and being younger than the preceding generation. For instance, he uses Labov's material from Martha's Vineyard (aw) that covers three generations; namely, Oldest Generation, Middle Genera-

tion, and Youngest Generation. But note that the three generations correspond to age level 61 to 90, age level 31 to 60, and age level 30 and under, respectively (cf. Labov 1972, 22 and 279), and that there is no information about the changes that the younger age groups will exhibit when they reach the older bracket. Thus, when the numerical values (Johnson 1976, 168) of 0.06, 0.37, and 0.88 are compared, the differences do not represent the acceleration of change rate in three succeeding generations; rather, they indicate three static manifestations of one continuous change taken cross-sectionally. In order to get the dynamics of change, what Johnson should have done would be something like this: Wait for the people of the younger generation to reach the next age level (i.e., 30 years) and then compare their centralization with that of the older generation at the same age level. For instance, he should have got the numeric value of the Youngest Generation (under 30) when they reach the next age level (31-60) and compare it with the numeric value of the Middle Generation when they are still at the level of 31-60 and do likewise for the Middle Generation and the Oldest Generation. But nothing of this sort has been done. Consequently, he has no data whatsoever to support the claim of acceleration in the rate of change.

By contrast, our data from the area study show exactly this kind of dynamics pertaining to change. That is, the results of all age groups investigated in 1950 were compared, 21 years later, with those of similar age groups investigated in 1971. Thus, we have information not only on two comparable age groups, say, 35-44, one taken from the 1950 study and the other from the 1971 study, for comparison pertaining to change, but also on different age groups taken cross-sectionally for comparison pertaining to the rate of change. The data from the area study are then backed up almost one to one by our data from the panel study. Thus, in the case of sound change, we can comfortably conclude that all age groups have changed but that the rate of change goes down as the age goes up within each generation.

From the aforementioned it must follow that there is a certain degree of incongruity in Johnson's data. For instance, how can he be sure that the first generation (Oldest Generation) did not have a faster rate of change when they were younger and that the third generation (Youngest Generation) will not slow down when they grow older? In fact, his data support precisely what we have found if

his three generations are regarded cross-sectionally, which is to say that the rate of change will be reduced in all cases, e.g., Martha's Vineyard, as one goes from the Youngest Generation (0.88) through the Middle Generation (0.37) to the Oldest Generation (0.06). The fact that Johnson has no data for each individual within one generation (which, by contrast, we have in the panel study) regarding his or her changes suggests that he cannot be sure of the rate of change being faster in each succeeding generation, that is, accelerating in succeeding generations. To demonstrate this fact, let me resort to a schematic representation of language change.

Figure 3 depicts sound change within one's own life time (notably from 15 to 44 years of age) with a plotted extension (dotted line) beyond 44. I have also circled three places which correspond to Labov's three age levels utilized in Johnson's data. The result of these modifications in the schematic representation is recapitulated as follows:

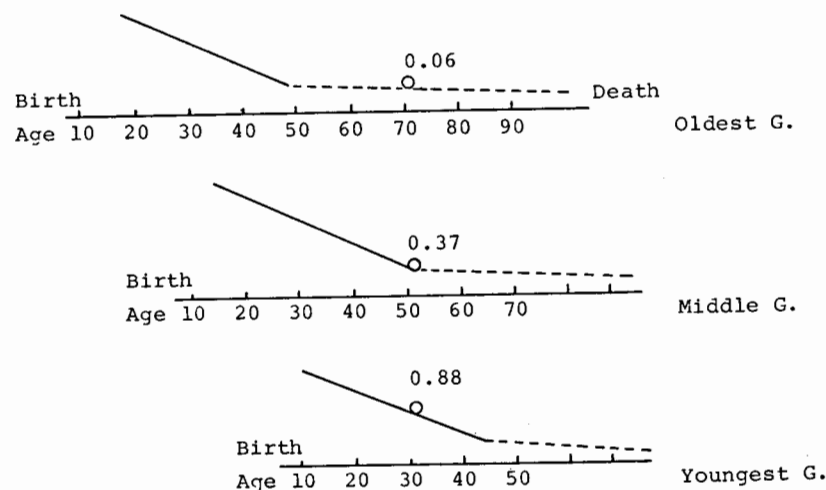


Figure 3  
Schematic representation of sound  
change and its rate

Observe now that this schematic representation shows that what Johnson has done is pick the three age brackets, one from each generation, with differing numeric values of vowel centralization (each of which falls in line with and can be explained by the rate of sound change therein). From my point of view, then, that the Oldest Generation has the lowest numeric value is not because, as Johnson has claimed, change begins slowly at first but because the age bracket (61-90) picked has, according to the schematic representation, already slowed down the rate of change; and likewise, that the Middle Generation and the Youngest Generation have successively increased their numeric values may also be explained by the fact that in the schematic representation they are younger in age and, therefore, stand higher in the rate of change. Consequently, it is not at all because change begins slowly and accelerates in succeeding generations, as claimed by Johnson who also thinks that he lends support to the claim of Wang and Cheng (in their discussion of lexical diffusion) and of Bailey that sound change follows an S-curve (Johnson 1976, 168). (By an S-curve is meant that sound change begins slowly and then increases rapidly [in some cases leaving residue].) In the light of my explanation above, it should be clear that none of the assertions made by Johnson and others is true.

At this point, I must add that certain sounds are more susceptible to changes than certain others (Peng 1976, 84 and 90). Given this view, which is supported by factual data from Japanese, the rate of sound change cannot be taken to mean that all language sounds (in a given language) progress in the same direction or at the same pace. Neither is it the case that the same type of sounds (in different languages) should have a fixed rate of change.

English may be cited as an example which shows marked ongoing changes in vowels rather than in consonants. This is, of course, historically true as well. However, another language, like Japanese, does not necessarily follow suit; my own study (Peng, 1976) clearly suggests that in Japanese consonants are much more susceptible to change. Thus, to answer Q3, I must say that whatever there is to discover regarding sound change in progress based, say, on English vowels, cannot and must not be generalized to apply to another language, unless there is a very good factual ground on which to build such a theoretical construct. I hope historical linguists have learned the lesson from the past, never to repeat the same mistake in the current exploration of sound change.

Conclusion

Let me now summarize by presenting three points, so as to bring the whole presentation to a close. Firstly, although linguists have been aware that when we speak of change it is people who change, and sound change is simply a manifestation (or symptom) of human change, not enough research is being done in, or attention paid to, the probe of what I have called the dynamics of change. This kind of study requires both cross-sectional and longitudinal investigations of fairly large samples in the same areas with the same method at an interval of hopefully 20 years. Since research of this nature is often painstaking and costly, historical linguists should turn to linguistic geographers and other social scientists for assistance in the provision of advice and materials; in spite of linguists like Kuryłowicz, who once renounced all support from linguistic geography and other social sciences for internal reconstruction (1964), it is through this kind of cross-fertilization that language scientists can hope to achieve the goal of dealing with the dynamics of change, among other things.

Secondly, I have also presented sufficient evidence to support my claim that if it is people who change, the change itself must take place within each individual to begin with, whose rate of change is affected by his or her own physical condition (age or maturation) as well as by the environment. Thus, as each individual increases his or her age, the rate of change decreases. Nobody knows, however, what will happen if life expectancy is extended beyond 100 years of age, to the rate of change.

Of course, life expectancy alone is not the influencing factor of human change; the environment counts heavily in this regard, the foremost influencing factor in the environment being human interaction. Note here that although the Japanese now live longer (perhaps longest?), 57% of the Japanese population is crowded on only 2% of the land, according to the latest report prepared by the Prime Minister's Office (The Japan Times, June 27, 1977). In this respect, then, Johnson is probably right in saying that "change proceeds more rapidly in urban than in rural areas" (1976, 165). I have reached a similar (albeit more substantial and elaborated) conclusion (Peng 1978).

Finally, I should mention that if human change is the key to sound change, more rigorous research is needed in such realms of

specialization as phonetics, neurolinguistics, sociolinguistics, and pedolinguistics to help determine the change and development in the total behaviors of humans as organisms.

References

- Johnson, Lawrence (1976): "A rate of change index for language", Language in Society 5, 165-172.
- Kuryłowicz, Jerzy (1964): "On the methods of internal reconstruction", in Proceedings of the Ninth International Congress of Linguists, 9-31, H.G. Lunt (ed.), The Hague: Mouton.
- Labov, William (1972): Sociolinguistic Patterns, Philadelphia: University of Pennsylvania Press.
- Nomoto, Kikuo et al. (1974): Chiiki Shakai no Gengo Seikatsu (Language behavior of a speech community), Report 52, Tokyo: National Language Research Institute.
- Peng, Fred C.C. (1976): "A new explanation of language change: The sociolinguistic approach", Forum Linguisticum 1, 67-94.
- Peng, Fred C.C. (1978): "Urbanization and language sciences: The Japanese case", in Language in Context, Fred C.C. Peng (ed.), Hiroshima: Bunka Hyoron Publishing Company.
- Peng, Fred C.C. (n.d.): "Sound change and language change: A sociolinguistic overview", special lecture delivered at the 1978 Annual Conference of the Linguistic Society of Japan (forthcoming in Language Sciences, 1978-9, vol. 1).

## TEMPORAL RELATIONS WITHIN SPEECH UNITS

## Summary of Moderator's Introduction

Ilse Lehiste, Department of Linguistics, Ohio State University

The title of the symposium leaves open the question of the type and size of the speech units. The contributors to the symposium have indeed chosen to address themselves to units of quite different types and sizes. Likewise, they have approached the problems connected with the temporal structure of speech units both from the perspective of speech production and from that of speech perception. The contributions include highly theoretical papers, papers presenting detailed results of experiments, and papers falling between these two poles. Some systematization appears to be in order. I would like to present herewith a framework within which I believe the issues can be profitably formulated for the discussions which I hope will follow.

The framework involves three dimensions. One of them concerns the relationship between timing control in production and the role of timing in perception. The second dimension deals with the direction of determination in the temporal organization of spoken language; specifically, with the question whether the timing of an utterance is determined by its syntax, or whether there exist rhythmic principles in production and perception that are at least partly independent of syntax. The third dimension follows directly from the previous two and relates to the type and size of speech units. What is the nature of those units, and are they to be established on the basis of a morphosyntactic analysis of the sentence, or on some kinds of independent phonetic criteria?

Clearly both production and perception are involved in oral communication by spoken language, and it would seem unnecessary to elaborate the point. However, I have had occasion to argue--against considerable weight of opinion--that durational differences in production, be they ever so significant statistically, cannot play a linguistically significant role if they are so small as to be below the perceptual threshold. It would be wise, I think, to remind oneself periodically of "the evident fact that we speak in order to be heard in order to be understood" (Jakobson et al. 1952). I hope, therefore, that in our discussion of temporal relations within speech units, models of production and models of perception will be related to each other.



The second and third questions concern the direction of determination: does phonology follow syntax, or are we dealing with interacting, but parallel hierarchies? Some researchers have developed programs for generating the temporal structure of a sentence on the basis of segments and syntactic structure, without paying any attention to rhythm. This is, I believe, due to a particular theoretical orientation. Generative phonology operates with segmental features; even suprasegmental features are attached to segments. And in a generative grammar, phonetic output is the last step in the generation of a sentence. An independent rhythm component simply has no place in the theory. For those scholars, then, the speech units are segments, phrases, clauses, and sentences. (And it is quite interesting to see them struggle with units not foreseen in the theory, like syllables and phonetic words.) Researchers who are not fully committed to this theoretical viewpoint operate with certain other units, such as speech measures or metric feet. Again, the reality of both kinds of units can be studied from the point of view of production as well as from that of perception.

Practically all the issues I have outlined are treated in the papers contributed to this symposium. Production is the main concern of the papers of Allen, Bannert, Klatt, and Öhman et al.; perception is the focus in the papers of Carlson et al., Donovan and Darwin, Fujisaki and Higuchi, Huggins, and Nootboom.

Among the papers dealing with production, Bannert considers the effect of sentence accent on the duration of VC sequences, employing a rather complex concept--vowel-to-sequence ratio  $V/(V+C)$ . The relationship between the VC-unit and its two parts represents a measure of the temporal structure of quantity of complementary length. Bannert shows that this unit is useful in describing the effect of the addition of sentence accent to quantity in Stockholm Swedish; it remains to be demonstrated whether the unit is as significant for perception as it is for production.

The paper by Klatt presents a detailed scheme for the synthesis by rule of segmental durations in English sentences. It is an almost pure example of the approach that starts from an abstract linguistic description and ends up as a sequence of segments, whose durations are conditioned by other segments and by syntactic con-

straints. Interestingly, a companion paper by Carlson et al. testing the output of Klatt's synthesis algorithm arrives at the conclusion that certain aspects of the durational pattern have greater perceptual importance than others. Vowel duration is more important than consonant duration; the durations between stressed vowel onsets seem to constitute a particularly important aspect of sentence structure.

The papers by Öhman et al. and by Allen concern themselves with production models in general. Öhman's et al. paper argues for a gesture theory of speech production. Their examples deal primarily with the assignment of fundamental frequency and are thus somewhat outside of the current topic. Allen's paper draws a useful distinction between descriptive models and theoretical models of speech timing, and makes the intriguing prediction that theoretical models may be about to undergo substantial modification, primarily due to the emergence of an "action theory" of speech production. According to that theory, neural activity is hierarchically organized into successively higher levels of coordination, until the highest level of all can only be described in terms of the overall goal of the action.

Among the papers devoted primarily to perception, Nootboom presents a decision strategy for the disambiguation of vowel length in Dutch. The strategy is complex, but listeners are fully capable of applying it in ongoing perception. Fujisaki and Higuchi present an analysis of the temporal organization of segmental features in Japanese disyllables consisting only of vowels, and find that although the onsets of the transition for the second vowel are distributed over a relatively wide range, a perceptual analysis of the onset of the second vowel shows relatively little temporal variation. It thus seems that the apparent diversity of the onset of transition in various disyllables is introduced to maintain the uniformity of perceived duration of segments. Fujisaki and Higuchi consider their results supportive of a model in which the motor commands and the articulatory/acoustic realizations of successive segments are programmed in such a way that the perceptual onsets of successive segments are isochronous.

The last two papers are likewise concerned with speech rhythm. Huggins finds that a correct rhythmic pattern, which is basically isochronous, enhances intelligibility, while a badly distorted

timing pattern impairs it seriously, even though all phonemes are identifiable. Donovan and Darwin deal with the perceived rhythm of speech, and give special consideration to the problem of isochrony. Their paper tests, among others, a hypothesis that I had formulated in 1973 and discussed in more detail in 1977. My observation was that listeners tend to hear utterances as more isochronous than they really are, and that listeners perform better in perceiving actual durational differences in non-speech as compared to speech. I concluded from this that isochrony is largely a perceptual phenomenon. Donovan and Darwin have confirmed these results. They make two points in addition: first, that isochrony is a perceptual phenomenon which is not independent of intonation, and second, that it is a perceptual phenomenon confined to language--reflecting underlying processes in speech production. Donovan and Darwin question the value of seeking direct links between syntax and segmental durations rather than indirect ones by way of an overall rhythmic structure.

I should like to propose a few direct questions for starting the discussion. What is the relationship between rhythm and syntax? How should rhythm be integrated into models of speech production and perception? What are the physiological constraints within which the production and perception of temporal structure must take place? What, indeed, is the nature of the temporal relations within speech units?

#### References

- Jakobson, R., C.G.M. Fant, and M. Halle (1952): Preliminaries to Speech Analysis, Cambridge, Massachusetts: MIT Press, tenth printing, 1972.
- Lehiste, I. (1973): "Rhythmic units and syntactic units in production and perception", JASA 54, 1228-1234.
- Lehiste, I. (1977): "Isochrony reconsidered", JPh 5, 253-263.

FORMAL AND STATISTICAL MODELS OF SPEECH TIMING: PAST, PRESENT,  
AND FUTURE

George D. Allen, Dental Research Center, University of North  
Carolina, Chapel Hill, N.C. 27514, USA

Let us begin this paper, whose goal is to review the kinds of models that have been developed in studies of timing in speech production and to suggest some possible directions for further research, by addressing briefly the general nature of models. Although the usual sense of the word "model" is that of an analogy, there is much room for differences in usage. On the one hand, we can have a "descriptive model", which models a set of observations, or data; a full-blown theory, on the other hand, models a complex and usually interacting set of constructs. Intermediate between these two extremes lies the single hypothesis, which is a projection of a theory onto a subspace of smaller dimensionality (often a single dimension) and which is "tested" by comparing it with a set of data. There are no theoretical boundaries between these three types of models, and most studies which model some aspect of speech timing contain elements of two (but seldom all three) of these categories.

Besides differing in the complexity of the structures which they reflect, models also differ in the intended accuracy with which they reflect those structures. Some models, for example, are intended primarily as conceptual guides, with only a loose fit between them and any existing data. Such models motivate the design of further studies and the analysis of data gathered by them, with the usually explicit goal of validation and refinement of the original model. The "chain" and "comb" models suggested by Bernstein (1967) were of this sort and have served as the basis for many recent studies of speech timing control. Other models are tailored closely to data or some other real world phenomena, their intent being more to parameterize the data (for example to permit comparisons of these parameters between different groups of speakers) than to explain the process whereby the data are generated. Klatt's (1975) study is an example of this kind of data-matching model. As with the descriptive vs theoretical distinction mentioned above, these extremes also allow much room for differences among models: about the only commonality among models in their "goodness-of-fit" to the data is that no model fits as well as its proponent would like.

A third difference among models is what I have chosen to term "formal" vs "statistical", though here again there is no true boundary between them. This contrast is exemplified by the difference between "regression" and "correlation" in statistics, the first being used to describe the form of the relationship between two measures (e.g., if A is 10 cm taller than B, then A may be expected to weigh about 5 kg more), the second an estimate of the strength of that relationship (e.g., A will weigh  $5 \pm 1.4$  kg more, 95 percent of the time). Lindblom and Rapp (1973) have thus in this sense developed a formal model of segment duration, whereas Kozhevnikov and Chistovich (1965) carried out the first of many statistical studies seeking significant negative correlations between the durations of successive segments as evidence of temporal compensation within production units.

Let us now review some past and present models of speech timing and its control in terms of these different general features. This review unfortunately cannot begin to cover the wealth of studies that now exist in this area. It would be useful, for example, to try to relate models of production to models of perception. Instead I shall restrict attention here to just a representative sample of models of timing in speech production and hope that perhaps the symposium itself will bring about the more complete discussion this topic deserves.

One major class of models has concerned the durations of segments, the earliest studies dealing with vowel duration in English (House and Fairbanks, 1953; Peterson and Lehiste, 1960; Kim, 1966). Although all of these were primarily descriptive in their origin, Kim's was the most theoretical in its intent. By explicitly labeling the branches on his tree with fixed durational values to be attributed to plus- vs minus-tense vocalic nuclei or plus- vs minus-voicing of the arresting consonant, he cut some of the ties his model had to the data which generated it and aligned it as well as he could with the constructs of distinctive feature theory.

More recent models of segment duration are those of Lindblom and Rapp (1973) and Klatt (1975), mentioned earlier. Both of these are descriptive models, though the data they describe, and thus their derivative models, are different. Lindblom and Rapp used phonologically restricted nonsense material and described variations in a segment's duration as a function of the number of seg-

ments, syllables, and words following it in the phrase. Klatt, on the other hand, used a meaningful paragraph, sacrificing control over word- and phrase-length comparisons while retaining contrasts in local segmental and prosodic context and adding syntactic contrasts. Interestingly, both of these studies describe segment variation as a contextually conditioned reduction in duration from a longest "base" form; several other related studies (e.g., Nootboom, 1972; Umeda and Coker, 1975) have done the same, and Keating and Kubaska (1978) have suggested a role for this process in speech development.

Although these carefully constructed models are in substantial agreement as to the major dimensions required for describing the durations of segments in the phonologically restricted speech samples from which they were derived, other investigators have suggested that they do not model "real" speech. Umeda and Coker (1975), for example, present an alternative model, based again on measured segment durations but from corpora that are less constrained by laboratory conditions than, say, Lindblom and Rapp's (1973) or Nootboom's (1972) data. Their data, and therefore their model, show the same local contextual effects as the others' (e.g., neighboring segment and syllable types, degree of stress, syntactic word classes), but the longer term effects (number of syllables remaining in the word, and words remaining in the phrase) are absent. This difference shows clearly one of the principle hazards associated with models derived from data: an apparently important component or dimension of the model may turn out to be an artifact of the observational situation. In this particular case the issue remains open.

There are many other studies of segment duration that deserve recognition here, and much more that might be said concerning those studies which have been mentioned. Because of space limitations, however, let us move on to a second major class of speech timing models, those which have dealt with the control of the articulatory time program. Aside from the oversimplified but heuristically useful "isochronic" model of English stress (cf., e.g., Pike, 1945), the first model of speech timing control appeared in Kozhevnikov and Chistovich (1965). As noted earlier theirs tried, via statistical techniques, to identify temporal compensation within production units, their underlying goal being to validate either the "chain" or "comb" model proposed by Bernstein (1967). Because of

procedural artifacts inherent to their method, however, they recognized that they could not decide the issues from their data, and so they abandoned the temporal domain in favor of the articulatory. Some later investigators (e.g. Lehiste, 1972; Wright, 1974) were not so cautious and claimed evidence for temporal compensation in spite of warning by Kozhevnikov and Chistovich (1965) and Ohala (1970) that variations in speech rate and measurement error could mask any true effects. Allen (1973, 1974), on the other hand, tried to circumvent the methodological problem by proposing a statistical model which used a statistic that was insensitive to rate variations and by including an explicit estimate of measurement error. In agreement with Ohala (1970, 1975) he found no evidence for temporal compensation within the freely spoken phrase, thus supporting the "comb" model (though only weakly, since a statistically negative result can never be strong evidence for any hypothesis).

In addition to examining the relative validity of the "chain" and "comb" models, Allen's model had the additional advantage of yielding a measure of the speaker's timing control accuracy. In one study (Cooper and Allen, 1977) this model was partially validated using speakers whose timing control was known to be poor, and in another (Tingley and Allen, 1975) the developing ability of children's speech timing control was charted. As a result of these limited successes, Allen (1978) suggested that the methodological limitations inherent in earlier statistical approaches to the study of speech timing control may yet be overcome.

Although Kozhevnikov and Chistovich (1965) and most other investigators were seeking to discover units of speech production, Allen (1973) was equally interested in determining the nature of the mechanism for speech timing control. Following Huggins (1972), Allen distinguished two possible models for such a mechanism ("capacitor discharge" vs "neural counter") and discussed evidence for and possible consequences of each. For example, although Creelman (1962) writes that his data are incompatible with any periodic clock for temporal discrimination, thus arguing against a cyclically activated neural generator, both Michon (1967) and Kristofferson (1976) present data with distinctly periodic components. No direct comparison of the various models suggested so far for controlling speech timing has been performed, however, and the issue remains open.

This brief sampling of models of speech timing may be summarized as follows. (1) Most studies modeling timing in speech production either have described the temporal properties of known production units, such as segments, or have sought evidence of unknown units or the mechanisms whereby they are produced. (2) Although there have been some methodological differences among studies, their results have been in substantial agreement, at least within major classes of models. (3) Many important issues raised by these studies are apparently testable, but great care will be required to avoid methodological pitfalls.

What is the shape of things to come in this area of study? Will tomorrow's models be refined variants of today's, or will new concepts force a radical restructuring of our thought? The answer, I believe, is "both". For some purposes, such as practical speech synthesis, refinements and straightforward extensions of present descriptive models will be adequate for some time. Here the output must be acceptable as fluent speech, but the process by which it is generated need not model human (neuro-) physiology.

There is already under way, however, at least one radical restructuring, which will affect profoundly the form of models of speech production and perception. Turvey (1975) and several of his colleagues have argued persuasively for what they call an "action theory" of speech production, in which the motor system's normal reflexes are organized into ever higher levels of coordination, the highest level of all being sensibly describable only in terms of the overall goal, or plan, of the action. Such mainstays of traditional speech production research as "segment", "coarticulation", and "motor unit" become, in this view, projections of the plan onto subspaces of greatly reduced dimensionality, so simplified in most cases as to obscure the "true" process of production. Fowler (1977) has examined the implications of this kind of theory for models of speech timing, giving us a good opportunity to glimpse at least the immediate future.

At a rather deep level of conceptualization, we may see more explicit appeal to the goals of the speech timing model; that is, it will be not only acceptable but even necessary to consider the function of temporal structure in order to understand adequately what we observe. For example, such a statement as "Speech is made to be spoken" (Allen, 1975) will become literal rather than

figurative truth.

Models of "intrinsic timing", as Fowler (1977) terms them, may impose far more explicit constraints on the domain of control than do many present-day models. Since in that view the temporal figure is as much a part of the speech act as, say, its neuromuscular features, intrinsic timing is an inherent property of the act, coterminous with it, not something that is imposed on it by an external timing generator that exists before and after as well as during. Hence it would be improper to speak, for example, of "the effect of speaking rate on segment duration", since the effect is really on the whole structured act within which the segment is embedded.

Some models already refer explicitly to domains of temporal constraint. Lindblom and Rapp (1973), for example, use one parameter to describe the effect of the number of syllables following within the same word and a second for the number of words following within the same phrase. Allen (1973) restricts his model of timing control to effects within the breath group. Other local constraints, such as neighboring phonemic context, are commonly imposed. Even so we may soon find the focus changing in our consideration of domains of temporal constraint; since the timing is intrinsic to the act, we would seek either to isolate acts as delimiters of temporal domains or to identify differences in timing control as evidence of action boundaries. We have often done this before, but usually intuitively or even unconsciously, and with segmental phonology and orthography as our guides. Following "action theory" into hierarchical systems of coordinated reflexes may bring us some interesting surprises.

Finally, we should still find as much need in models of intrinsic timing as in our present models for the notions of "temporal compensation" and "timing control mechanism" ("clock"). The assumption that motor action plans are organized hierarchically implies that temporal compensation will appear at all levels below the very highest; otherwise the temporal figure could not be intrinsic to the plan. Moreover, as long as neuromuscular events within the plan do not follow rapidly one upon the other, as fast as the associated lowest level reflexes allow, a controlling mechanism must be assumed to decide when to move on to the next. It could be proposed that the neural structures and pathways responsible

for coordinating the action of muscles in space are simultaneously responsible for their temporal patterning as well, i.e., the plan is its own clock. The dissociation of temporal from spatial control in such dysrhythmic conditions as cerebellar ataxia, however, suggests strongly that a separate mechanism will continue to be needed in adequate models of timing in motor action plans.

In conclusion, we may expect that descriptive models of speech timing will continue to be elaborated, with fairly clear lines of historical development from the very earliest descriptions of segment durations. Theoretical models, on the other hand, may be about to undergo substantial modification, as we revise our conceptualization of the speech production process and of the relationship of timing to that process.<sup>1</sup>

#### References

- Allen, G.D. (1973): "Segmental timing control in speech production", *JPh* 1, 219-237.
- Allen, G.D. (1974): "Measurement error in speech timing studies", *JASA* 55, Suppl. 1, S42 (abstract).
- Allen, G.D. (1975): "Speech rhythm: its relation to performance universals and articulatory timing", *JPh* 3, 75-86.
- Allen, G.D. (1978): "Vowel duration measurement: A reliability study", *JASA* 63, 1176-1185.
- Bernstein, N.A. (1967): *The Coordination and Regulation of Movements*, Oxford: Pergamon Press.
- Cooper, M.H. and G.D. Allen (1977): "Speech timing control in normal speakers and stutterers", *JSHR* 20, 55-71.
- Creelman, C.D. (1962): "Human discrimination of auditory duration", *JASA* 34, 582-593.
- Fowler, C.A. (1977): *Timing Control in Speech Production*, Indiana: University Linguistics Club.
- House, A.S. and G. Fairbanks (1953): "Influence of consonant environment upon the secondary acoustic characteristics of vowels", *JASA* 25, 105-113.
- Huggins, A.W.F. (1972): "Just noticeable differences in segment duration in speech", *JASA* 51, 1270-1278.
- Keating, P. and C. Kubaska (1978): "Variation in the duration of words", *JASA* 63, Suppl. 1, S56 (abstract).
- Kim, C.-W. (1966): "The linguistic specification of speech", UCLA: Working Papers in Phonetics, 5.
- Klatt, D.H. (1975): "Vowel lengthening is syntactically determined in a connected discourse", *JPh* 3, 129-140.

---

(1) The support of NSF grant number BNS 7614345 and NIH grant number RR 05333-16 is gratefully acknowledged.

- Kozhevnikov, V.A. and L.A. Chistovich (1965): Speech: Articulation and perception, Joint Publications Research Service, Washington, D.C., 30,543.
- Kristofferson, A.B. (1976): "Low-variance stimulus-response latencies: Deterministic internal delays?", Perc.Psych. 20, 89-100.
- Lehiste, I. (1972): "Timing of utterances and linguistic boundaries", JASA 51, 2018-2024.
- Lindblom, B. and K. Rapp (1973): "Some temporal regularities of spoken Swedish", Papers from the Institute of Linguistics, University of Stockholm.
- Michon, J.A. (1967): Timing in Temporal Tracking, Soesterberg, the Netherlands: Institute for Perception.
- Nooteboom, S.G. (1972): Production and Perception of Vowel Duration, Doctoral dissertation, University of Utrecht.
- Ohala, J.J. (1970): "Aspects of the control and production of speech", UCLA Working Papers in Phonetics 15.
- Ohala, J.J. (1975): "The temporal regulation of speech", in Auditory Analysis and Perception of Speech, G. Fant and M.A.A. Tatham (eds.), New York: Academic Press, 431-452.
- Peterson, G.E. and I. Lehiste (1960): "Duration of syllable nuclei in English", JASA 32, 693-703.
- Pike, K.L. (1945): The Intonation of American English, Ann Arbor: The University of Michigan Press.
- Tingley, B.M. and G.D. Allen (1975): "Development of speech timing control in children", Child Devel. 46, 186-194.
- Turvey, M.T. (1975): "Preliminaries to a theory of action with reference to vision", in Perceiving, Acting and Knowing: Toward an Ecological Psychology, R. Shaw and J. Bransford (eds.), Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Umeda, N. and C.H. Coker (1975): "Subphonemic details in American English", in Auditory Analysis and Perception of Speech, G. Fant and M.A.A. Tatham (eds.), 539-564, New York: Academic Press.
- Wright, T.W. (1974): "Temporal interactions within a phrase and sentence context", JASA 56, 1258-1265.

## THE EFFECT OF SENTENCE ACCENT ON QUANTITY

Robert Bannert, Phonetics Laboratory, Lund University, Sweden

This paper will focus on the phonological aspects of the temporal structure of quantity in Central Swedish. Its domain is the vowel and consonant sequence resulting in the temporal pattern of complementary length, namely vowel + short consonant (V:C) or short vowel + long consonant (VC:). Quantity will be studied from the sentence perspective by investigating the prosodic effect of sentence accent (SA) on the VC-sequences.

Investigation

The tonal manifestation of SA in Stockholm Swedish is treated exhaustively by Bruce (1977). The present investigation builds on his tonal findings. Two speakers from Stockholm, EH, female, the main informant in Bruce (1977) and TB, male, representing the same dialectal variety, produced the test material. The test words were stöka (V:C) and stöcka (VC:). The qualitative difference between the long and short vowel is rather small. They were placed alternatively in one of the three positions (1,2,3) in the base sentence "Man kan lämna långa nunnor efter åtta." (One can leave tall nuns after eight). The test words, like the basic words (underlined) in the three sentence positions, are stressed with word accent 2 on the first syllable.

The test material consisted of 12 sentences. Six of them contained the test words in one of the three accented positions without SA, SA falling on the adverbial. The other six sentences contained the test words with SA.

Results and discussion

The VC-sequences without sentence accent are regarded as a reference for studying the effect of SA on quantity.

1. The temporal structure of quantity. The overall means of the durations of the vowel and the following consonant are plotted in figure 1. The diagram represents the temporal space for the manifestation of the VC-sequences (lower limitations in terms of incompressibility, cf Lindblom et al 1976, left aside). The range is illustrated by ellipses, constructed according to the standard deviations of vowel and consonant. The data points of both speakers and both types of sequences fall in such a way that they may easily be accounted for by straight lines. The temporal structure



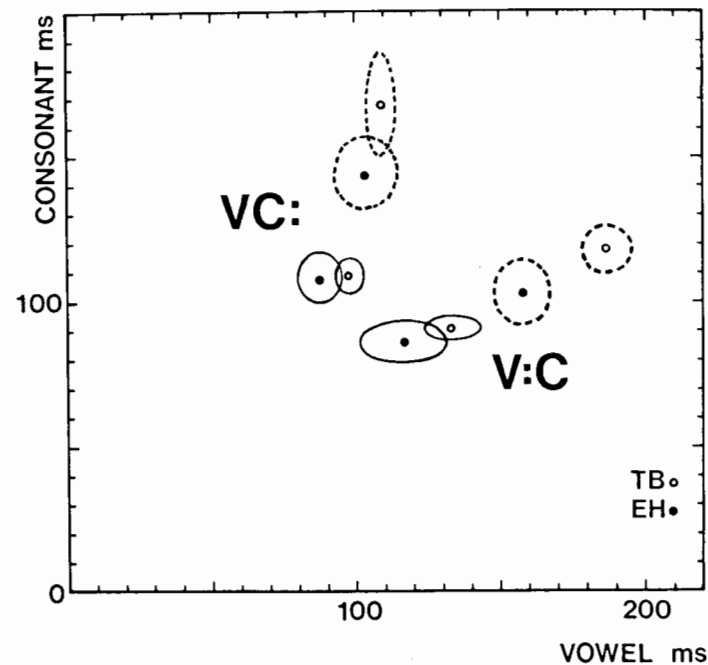


Figure 1. The effect of sentence accent on the temporal structure of the /V:C/- and /VC:/-sequences, consonant durations plotted against vowel durations. Sentence accent with broken-line circles. Pooled means and standard deviations for two speakers from Stockholm.

of the two types of sequences with and without SA are very similar for both speakers. Figure 1 shows also that the sequences with SA not only lie further apart from those without SA: Each type of sequence is also more separated from its counterpart.

The temporal effect of SA on quantity in Stockholm Swedish is a considerable increase of the segment durations which makes the temporal structure of the two contrasting VC-sequences more dissimilar. Thus the temporal contrast becomes clearer in focus position, given more prominence by the SA.

2. The increase of segment durations. The durations of the segments do not increase in a uniform way. It is evident from figure 1 that the increase of segment duration is largest for the long segment of each type of sequence, i.e. the long vowel in (V:C) and the long consonant in (VC:).

The V/C relations. The proportion of the increase of segment duration in each position and in all positions together is given in table 1. The ratio of the durational increase is defined as the relationship between the consonant and the vowel,  $k = \Delta C / \Delta V$ . It expresses the degree of lengthening of the consonant compared to that of the preceding vowel.

Table 1. Increase in segment durations. The factor  $k$  gives the proportion of the lengthening of the consonant compared to that of the preceding vowel.

sequence	speaker	position				
		1	2	3	1-3	pooled
V:C	TB	0.48	0.43	0.72	0.53	0.48
	EH	0.46	0.48	0.39	0.43	
VC:	TB	4.70	5.17	6.18	5.36	3.50
	EH	3.00	1.76	2.06	2.23	

Although there is some variation across the three sentence positions for both speakers, the short consonant is prolonged approximately by a factor of 0.5 of the preceding long vowel, while the duration of the long consonant increases much more in comparison with its preceding short vowel.

A non-uniform change of segment durations in VC-sequences with complementary length is also found in other prosodic contexts in Central Swedish and also in Central Bavarian, namely with differing speaking rates and stresses (contrastive vs neutral) and for words pronounced in isolation vs embedded in a sentence (cf Gårding et al 1975, Bannert 1976). This would suggest that the increase in segment duration due to different conditions is both linear and non-uniform.

The degree of lengthening. When the relationship of the non-uniform increase in segment duration between the vowel and the following consonant is established, it is sufficient to know the degree of lengthening for only one segment, e.g. the vowel. The relative degree of lengthening of the long and short vowel is given in table 2.

For this parameter, too, the change is not invariant. Both speakers behave differently for the three sentence positions and for the two categories of vowels. These differences may be accounted for with reference to two individual differences:

1) There are different durations between the speakers on the three

Table 2. Percentage of the increase in vowel duration with SA.

sequence speaker		position				
		1	2	3	1-3	pooled
V:C	TB	42.7	48.4	29.9	39.9	37.2
	EH	28.4	38.9	36.1	34.3	
V C:	TB	10.2	12.5	10.9	11.2	14.5
	EH	14.9	20.5	19.8	18.3	

sentence positions in the reference cases without SA (figure 1).

- 2) The speakers have two different tonal behaviours. Speaker EH has a tonal rise in the postconsonantal vowel, the pitch level of which is lower than that in the accented vowel. Speaker TB, however, reaches the high  $F_0$ -level for the SA, which is higher than that in the accented vowel, in the consonant itself. Therefore, TB seems to need more time for producing the consonant.

But these minor differences in temporal behaviour remain well within the typical pattern of complementary length.

### 3. Preserving the temporal identity of the VC-sequences

It can be hypothesized that the change in the temporal patterns of the VC-sequences is governed by a dominant phonological principle, which is due to perceptual mechanisms: Preserve or sharpen the temporal contrast.

Two ways of increasing segment durations. There are two possibilities for the means by which segment durations in focus position may be increased:

- (1) equally to both segments either in absolute terms (ms) or as a percentage of the segment duration without SA,
- (2) differently to both segments. Then the question arises: How different and why?

In figure 2, two kinds of equal increase of segment duration and their effect on the temporal pattern of the sequences are illustrated. The reference points are the pooled means for the durations of the /V:C/- and the /VC:/-sequences without SA for both speakers taken together.

An absolute increase in the duration in ms, equal for both the vowel and the consonant, will result in shifting the VC-points along a straight line, corresponding to the slope  $k=1$ . As a consequence, however, the temporal structures of the two contrasting sequences will become more similar to each other. It is known

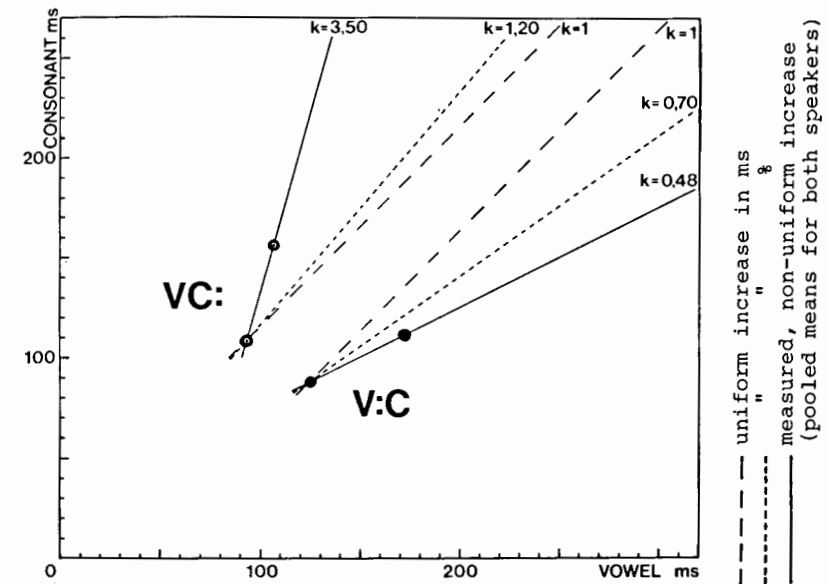


Figure 2. Different ways of increasing segment durations in the VC-sequences. The factor  $k$  gives the degree of lengthening for the consonant compared to the preceding vowel.

that durational differences must become greater with increasing segment duration in order to be perceived (cf the discussion in Lehiste 1970).

A relative increase of the duration, equal for both segments, is shown by the dotted lines. The slope (factor  $k$ ) is now smaller for the /V:C/-sequence and larger for the /VC:/-sequence due to the asymmetric temporal structure within the sequences. This relative increase will result in enlarging the temporal difference between the sequences. But it seems obvious that this equal relative change of duration does not lead to a sufficient dissimilarity between the sequences with SA, either. The measured pooled means clearly lie further apart from each other than either of the possibilities would predict. Whereas the duration of the long vowel increases twice as much as that of the following short consonant ( $k \approx 0.5$ ), the duration of the long consonant increases by far faster than that of the short vowel ( $k \approx 3$ ).

The segment-to-sequence ratio. Due to the temporal patterns of the VC-sequences, they can be viewed as a unit of production and per-

ception. Then the relationship between the unit and its two parts will represent a measure of the temporal structure of quantity of complementary length. The vowel-to-sequence ratio  $V/(V+C)$ , expressing this relationship, was introduced by Bannert (1976) and applied to three languages with complementary length, Central Bavarian, Northern Icelandic, and Central Swedish.

Because of the phonological dependencies between the vowel and the following consonant, it seems inadequate to state the temporal structure of VC-sequences with complementary length by calculating the segment ratios  $V/V:$  and  $C/C:$  (cf the criticism in e.g. Lindblom et al 1976). Neither of these ratios can account for the temporal changes of the two speakers.

The segment-to-sequence relations, given as the  $V/(V+C)$  ratios, are plotted in figure 3. It is clear that, when represented in this way, both speakers change the temporal structure of the sequences in exactly the same way.

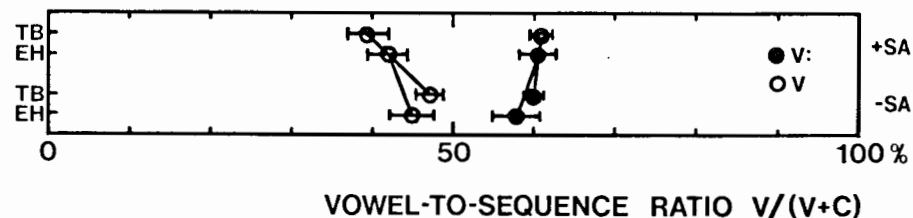


Figure 3. Variation of internal temporal structure of the VC-sequences in sentence accent position for each speaker. The relationship between the segments within the sequences is expressed by the vowel-to-sequences ratio.

The addition of SA to quantity in Stockholm Swedish increases the temporal distance, e.g. in terms of the segment-to-sequence ratio, between the two types of VC-sequences. Thus the temporal contrast for perception is well maintained or even enlarged.

#### References

- Bannert, R. (1976): Mittelbairische Phonologie auf akustischer und perzeptorischer Grundlage. Travaux de l'Institut de Linguistique de Lund X, B. Malmberg and K.Hadding (eds), Lund: Gleerup.
- Bruce, G. (1977): Swedish word accents in sentence perspective. Travaux de l'Institut de Linguistique de Lund XII, B.Malmberg and K.Hadding (eds), Lund: Gleerup.
- Gårding, E., O. Fujimura, H. Hirose, and Z. Simada (1975): Laryngeal control of Swedish word accent. Working Papers 10, 53-82. Phonetics Laboratory and Department of General Linguistics, Lund University.
- Lehiste, I. (1970): Suprasegmentals, Cambridge, Massachusetts: MIT Press.
- Lindblom, B., B. Lyberg, and K. Holmgren (1976): Durational patterns of Swedish phonology: Do they reflect short-term memory processes? Mimeographed. Department of Phonetics, Stockholm University.

## SOME NOTES ON THE PERCEPTION OF TEMPORAL PATTERNS IN SPEECH

Rolf Carlson\*, Björn Granström\*, and Dennis H. Klatt, Mass. Inst. of Tech., Cambridge, MA 02139 USA. [\*Also Dept. of Speech Communication, KTH, S-10044 Stockholm, Sweden.]

Introduction. Prosodic factors in speech have recently attracted a remarkable amount of linguistic and phonetic research. A prevalent point of view is that prosody is of paramount importance, both for naturalness and intelligibility of speech. As a result of this belief, a change can now be seen in the methods adopted in speech training for hard-of-hearing and foreign language students. The increased focus on suprasegmental compared to segmental articulation is possibly advantageous. From a scientific point of view, however, very little evidence is yet available on the quantitative importance of prosody. This is especially true of the relative importance of different aspects of the prosodic pattern.

From a study employing synthetic speech (Huggins, 1976), we know that really deviant durations and fundamental frequency contour decreases intelligibility. Prosodic parameters have also been shown to be effective in disambiguating sentences (Lehiste *et al.*, 1976). Our concern, however, has more to do with what information an explicit description of prosody has to supply and the precision with which it is supplied.

Descriptive models for segmental duration and fundamental frequency have been designed for a number of languages. Typically these models are based on material read repeatedly by a single speaker in a neutral, non-emphatic way. Subjects can perform remarkably consistently within such a recording session, but an examination of spontaneous speech reveals great variability in the prosodic realizations of a given sentence.

Thus it is not clear how precise the specification of duration is in the speech code common to speaker and listener. We also know that perception imposes certain restrictions on how prosodic effects could be appreciated (Klatt and Cooper, 1975). From previous studies (Carlson and Granström, 1975; Fujisaki, 1975), we know that the sensitivity to durational changes is greater in vowels than in consonants. The durational balance

between syllable nuclei, as well as the interval between onsets of stressed vowels (a measure related to the foot concept) have been shown to be perceptually important (Carlson and Granström, 1975; Huggins, 1972; Lehiste, 1977).

This leads to the questions that we wish to address: given a primary interest in the functional properties of a model of prosody, what demands should we put on it? What aspects of the description are most important? Will different models be ranked in the same order if different criteria such as naturalness and intelligibility are used?

In our present study we have evaluated both the naturalness and intelligibility of sentences with several different durational structures. As a starting point we have used a version of Klatt's durational rules for American English (Klatt, 1979) that we use in the MIT text-to-speech project (Allen, 1976).

Test Material. An algorithmically complete rule system is meant to generate a first order approximation to the durational structure of any spoken English sentence. In order to evaluate such a system of rules, a variety of syntactic and phonological structures ought to be tested. The test material in our experiments, presented below, could include only a small sample of such structures. These include the active, passive, question, simple, compound, and complex embedded sentence types. Both short and long noun phrases are represented. In Sentence 8, the ")n" after "seafood" is specially used to indicate that the following prepositional phrase is a sentential modifier, rather than modifying the "icy seafood" noun phrase.

Test Sentence	measured dur. (msec)	synth.dur. (msec)
1. Someone at the table )n ordered hot and sour soup.	2365	2615
2. Going to school )n was an adventure.	1625	1860
3. He who eats too much )c will become fat.	2105	2295
4. If Kate )n goes, Bill )n will eat her orange.	2430	2385
5. Old eggs )n often spoil french bread.	2200	2330
6. The fat brown turkey )n was chased by everyone.	2495	2415
7. Do you think that it will rain?	1370	1450
8. Pete )n ate icy seafood )n on the veranda.	2195	2155
9. Frank )n saw pretty streetcars )n in San Francisco.	2755	2845
where: Noun Phrase Boundary = )n, Clause Boundary = )c		

These sentences were recorded several times, the most natural sounding recording was selected, and the duration of each segment was measured. Since the rules are intended to match the speech of a particular speaker (DHK), the same subject was employed in the recording session. Nine different versions of each sentence were synthesized and put on language master cards. The synthesis algorithm is discussed in Klatt (1979). The versions listed below include three (3-5) that might be expected to be preferred over version Rule (since the Rule durations are adjusted in part toward Ref durations) and four versions (6-9) expected to be worse than Rule (since various rules contained in Rule have been deleted).

- 1 Ref            Synthesis by rule using the measured durations from natural speech, but normalized linearly over the whole sentence to get the same total duration as Rule. This adjustment of the speech rate was rather small, averaging 6 percent.
- 2 Rule           Synthesis using the rule system.
- 3 Vowel          Synthesis using the vowel durations from Ref and the consonant durations from Rule.
- 4 Cons           Synthesis using the consonant durations from Ref and the vowel durations from Rule.
- 5 StressVO      Synthesis using the durations from Rule but linearly normalized between stressed vowel onsets to get the same durations between onsets of stressed vowels as in Ref.
- 6 NoParse       Synthesis using the rule system, but disregarding the syntactic boundaries marked by ")n" and ")c".
- 7 SimpleFL      Same as Simple (below) but with clause-final lengthening at punctuation marks. Each segment after and including the last stressed vowel is assigned increased duration by a factor of 1.65.
- 8 Simple        Synthesis using a very simple rule system:  
                  Stressed vowel        : Dur= .80 \* inherent duration  
                  Unstressed vowel     : Dur= .60 \* inherent duration  
                  Stressed consonant   : Dur= .90 \* inherent duration  
                  Unstressed consonant : Dur= .65 \* inherent duration
- 9 Random        Synthesis using the reference duration, but randomly multiplied or divided by a factor determined by the deviation between Rule and Ref. This condition was included as a clear example of a bad system.

Experiment I: Naturalness. Nine phonetically trained subjects (native speakers of American English, working at RLE, MIT) were asked to sort the nine versions of each sentence according to naturalness of the durational structure. The subjects used a language master and headphones. After the order for a particular

sentence type was settled, the subject assigned a number corresponding to subjective naturalness (from 0 - 100) to each version. Most of the subjects finished the task within two hours. In some cases, the task required several sessions. Since the subjects used different scales in the rating task, the data from each subject were normalized to produce a mean of 0 and a standard deviation of 100. Mean ratings across sentences (Table 1, Column labeled "mean") indicate that Ref, Rule, Vowel, and StressVO are judged to be significantly more natural than the others.

The reproducibility of the naturalness rating for a subject was estimated from sentence seven, which had no syntactic markers in the input representation, and was thus identical in versions Rule and NoParse. The mean normalized distance across subjects for this pair was 26, which compares favorably to a typical standard deviation of 60 for the observations underlying an element in the matrix (Table 1). This suggests that subjects were quite consistent in their ratings compared to the intersubject variability. The estimated standard deviation of each element in the matrix is about 20 (60 divided by the square root of 9). The standard deviation of the mean across sentences is given in the table for each version.

Table 1. Naturalness ratings from Experiment I, as averaged across nine subjects. Column A indicates the number of errors for 6 versions used in an intelligibility test described in Experiment II. Versions 3, 4 and 9 were not included in Experiment II.

ver- sion	sentence									mean	st.d.	A
	1	2	3	4	5	6	7	8	9			
1	78	58	17	33	21	68	-18	51	78	43	8	2
2	18	33	47	32	70	70	86	17	23	44	7	12
3	86	60	28	0	52	104	44	81	40	55	7	-
4	0	-2	6	73	38	71	-25	-23	61	22	8	-
5	99	28	-2	81	35	108	-25	50	92	52	8	9
6	14	66	3	30	-101	33	70	11	54	20	8	12
7	-36	4	22	21	-114	4	43	6	-80	-14	9	15
8	-63	-66	3	-169	-146	-30	-125	13	-127	-79	11	24
9	-116	-168	-169	-132	-163	-148	-150	-190	-80	-146	9	-

Experiment II: Intelligibility. Some of the versions used in Experiment I were included in an intelligibility test that was presented individually to 18 MIT students. These subjects were phonetically naive, native speakers of American English, and unfamiliar with synthetic speech. Before the test was run, the subject listened to a short passage of synthetic speech (75 sec)

to get acquainted to the speech quality. This familiarization process has been shown to be very rapid (Carlson et al., 1976). The number of word errors out of 122 possible words (excluding articles) is shown in Column A of Table 1, and is plotted against the naturalness rating data in Fig. 1.

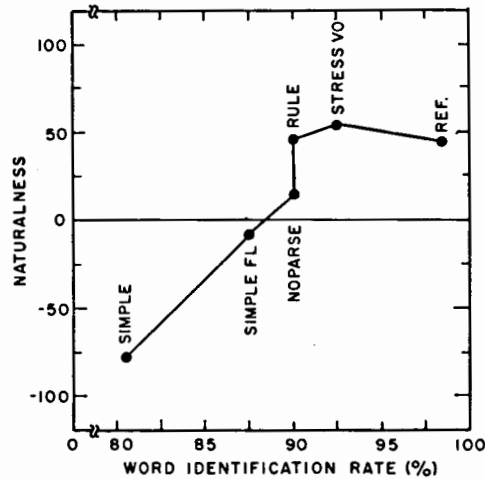


Figure 1. Mean naturalness ratings for six versions are plotted against the word identification rate from Table 1. The two measures are positively correlated, but the improvement in naturalness from NoParse to Rule is not accompanied by an improvement in intelligibility. It should be emphasized that the intelligibility figures are based on a small amount of data and should be interpreted with some caution.

**Discussion.** It is clear from ratings and comments given on the answer sheets, that subjects have different preferences. For example, Ref is not considered best by all subjects for all sentences. This might be a question of dialectal preference or idiosyncratic differences. Another possibility is that durations from natural speech, imposed on synthetic speech with a somewhat different realization of F0 and segmental content could constitute an incompatible combination. There is no way of controlling for this in the present study. A parallel study using LPC-coded natural speech might shed some light on this issue.

Ref and Rule have about the same mean naturalness score in Table 1, indicating that the durational rules produce as natural a durational structure as our reference speaker <POINT 1>. However, it should be noted that the test material consist of rather short sentences without e.g. the semantic relations between sentences that exist in paragraph-length material and that the intelligibility of Rule was somewhat lower than that of Ref.

We wanted to examine how the intermediate versions between Ref and Rule (Vowel, Cons, and StressVO) are ordered. This could not be done if we are not sure that the relation between Ref and Rule is the same for all subjects. Therefore, in Table 2, the results are presented after discarding the data on a sentence for each subject who rated Rule higher than Ref. (This will, of course, reduce the naturalness score for Rule relative to all other versions.)

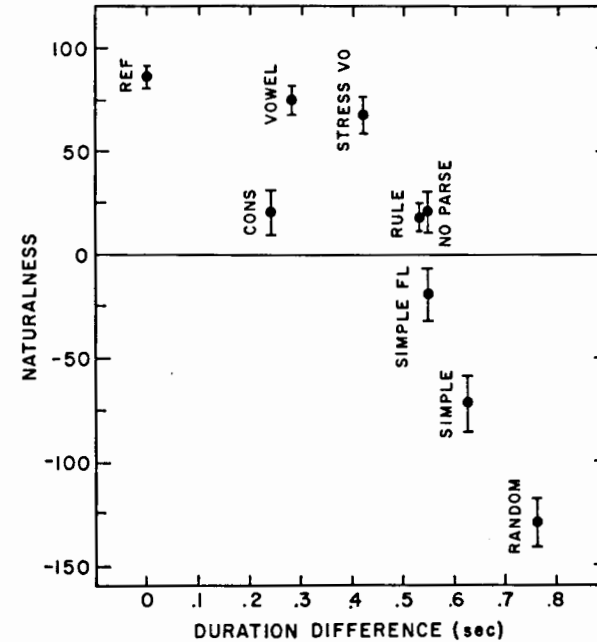


Figure 2. Mean naturalness ratings from Table 2 are plotted against one measure of the physical durational distance to Ref (city block), i.e., the sum, over all segments, of the absolute difference in duration between the version and Ref (average per sentence). There is a general correlation between naturalness ratings and physical difference in duration, but Rule, StressVO and Vowel are rated more natural than one might expect given the durational differences involved.

Table 2. Naturalness ratings after excluding Ss who rated Rule better than Ref. The number of subjects for each sentence is marked in the last line.

ver- sion	sentence									mean	st.d.
	1	2	3	4	5	6	7	8	9		
1	102	85	66	55	73	102	93	96	83	85	6
2	13	12	10	4	59	42	44	14	12	18	8
3	110	85	78	6	73	116	72	82	40	74	8
4	2	-20	-42	92	90	91	25	-43	52	20	11
5	105	27	27	53	42	116	41	56	90	67	9
6	-4	43	-30	52	-125	37	44	19	56	20	10
7	-40	-10	36	26	-159	-11	17	13	-82	-20	13
8	-42	-68	-27	-176	-208	-30	-71	-4	-119	-73	14
9	-125	-133	-150	-121	-122	-150	-116	-180	-75	-130	12
	7	6	5	5	2	5	2	7	8		(# subjects)

The most striking result seen in Figure 2 is that both Vowel and StressVO are significantly more natural than Cons, despite their greater durational distance from Ref. This corroborates earlier observations that these two durational units i.e. vowel duration and interval between onset of stressed vowels are of great perceptual importance <POINT 2>. Cons, which has all consonant durations right but vowel durations done by rule, does not score significantly better than Rule, reinforcing this interpretation. Furthermore it is obvious that physical distance is clearly not a reliable predictor of perceptual distance.

Isochrony, i.e. the tendency toward equal durations between certain units, has been discussed in the literature. It might be suspected that the high scores for Ref and StressVO are because they preserve the isochrony of real speech. We compared Rule and Ref to see which has a greater tendency toward equal distances between stressed vowel onsets. If anything, Rule is more isochronous than Ref, suggesting that the amount of isochrony implemented in the rules via, e.g., cluster shortening and unstressed segment shortening is probably sufficient, and no "isochrony rule" per se need to be added <POINT 3>.

Versions Rule and NoParse have the same naturalness score in Figure 2. However, it must be remembered that an editing has taken place which selectively lowers the score of Rule. In Table 1, these two versions are significantly different. For some sentences, however, the score for NoParse is higher than that for Rule. Even if these differences are not highly significant, they indicate that in these instances, NoParse is regarded as close in quality to Rule. One possible reason could be that the rules dealing with phrase final lengthening overexaggerate the lengthening effect. An analysis yielded no support for this interpretation in our data. Another possibility which seems more reasonable is that the phrase-final lengthening rule is applied too frequently <POINT 4>. A simple-minded cure might be to ensure that short phrases containing only one content word are not affected by phrase final lengthening although recent work by Cooper et al (1978) indicates the likelihood of a more complex relation between surface structure and lengthening.

Comparing Simple and Rule, we can conclude that rules modifying the duration of a segment as a function of syntax and segmental context are of significant importance for both naturalness and intelligibility. Approximately half of the difference between the two versions seems to be explained by the extremely simple clause final lengthening rule used for SimpleFL <POINT 5>.

The intelligibility results shown in Figure 1 indicate a clear correlation between intelligibility and naturalness. Correct durations result in significantly better intelligibility and naturalness. This confirms in part the current belief in the importance of prosody to sentence perception. <POINT 6>.

#### References

- Allen, J. (1976), "Synthesis of Speech from Unrestricted Text", *Proc. IEEE* 64, 433-442.
- Carlson, R. and Granström, B. (1975), "Perception of Segmental Duration", in *Structure and Process in Speech Perception*, A. Cohen and S.G. Nooteboom (Eds.), Springer-Verlag, Berlin, 90-104.
- Carlson, R., Granström, B., and Larsson, K. (1976), "Evaluation of a Text-to-Speech System as a Reading Machine for the Blind", *STL QPSR* 2-3/1976, 9-13.
- Cooper, W.E., Paccia, J.M., and Lapointe, S.G. (1978), "Hierarchical Coding in Speech Timing", *Cognitive Psychology* 10, 154-177.
- Fujisaki, H., Nakamura, K., and Imoto, T. (1975), "Auditory Perception of Duration of Speech and Non-speech Stimuli" in *Auditory analysis and perception of speech*, G. Fant and M. Tatham (Eds.), Academic Press, London.
- Huggins, A.W.F. (1972), "On the Perception of Temporal Phenomena in Speech", *J. Acoust. Soc. Am.* 51, 1279-1290.
- Huggins, A.W.F. (1976), "Speech Timing and Intelligibility", *Proc. Attention and Performance* 7, J. Reguin (Ed.).
- Klatt, D.H. (1979), "Synthesis of Segmental Durations in English Sentences", *9th International Congress of Phonetic Sciences*, Copenhagen.
- Klatt, D.H. and Cooper, W.A. (1975), "Perception of Segment Duration in Sentence Contexts", in *Structure and Process in Speech Perception*, A. Cohen and S.G. Nooteboom (Eds.), Springer-Verlag: Heidelberg.
- Lehiste, I. (1977), "Isochrony Reconsidered", *J. Phonetics* 5, 253-263.
- Lehiste, I., Olive, J.P., Streeter, L.A. (1976), "The Role of Duration in Disambiguating Syntactically Ambiguous Sentences", *J. Acoust. Soc. Am.* 60, 1199-1202.



## THE PERCEIVED RHYTHM OF SPEECH

Andrew Donovan and C.J. Darwin, Laboratory of Experimental Psychology, and Centre for Research in Perception and Cognition, University of Sussex, Brighton, England

Introduction

Attempts to model the duration or rhythm of the segments of connected speech fall into two broad groups (see Fowler, 1977, for a review). On the one hand are those which allow the syntactic structure of the utterance to perturb a segment's ideal duration (e.g. Lindblom & Rapp, 1973; Klatt, 1975) but which recognise no overall rhythmic patterning; on the other hand are those which allow an overall rhythmic pattern to be perturbed by limits on the segmental durations which must be compressed or expanded into that pattern (Abercrombie, 1964; Witten, 1977). This latter family of models has taken for its underlying rhythm a sequence of isochronous beats occurring on adjacent stressed syllables marking out rhythmic units called feet. The choice of an isochronous foot is based partly on linguistic intuition ("English utterances may be considered as being divided by the isochronous beat of the stress pulse into feet of (approximately) equal length". Abercrombie, 1964), and partly on the observable phonetic fact that syllables tend to be shorter, the more there are in a foot (Huggins, 1975; Fowler, 1977). In choosing between these two approaches a crucial question is the status of the isochronous beat. It is clearly not a phonetic fact since a foot with many syllables tends to be longer than one with fewer (Halliday, 1967; Allen, 1975). Is isochrony then a significant linguistic insight, or merely a poetic fiction? Lehiste (1973; 1977) has argued that the discrepancy between the linguistic intuition and the phonetic data may be due to a perceptual illusion. Perhaps we hear speech as more isochronous than it actually is. Indeed such an illusion is precisely what we would expect if perception undid those perturbations required by segmental constraints on an underlying regular rhythm, presenting to the listener the underlying rhythm of the speaker. Evidence for the perceptual reality of isochrony would thus argue for its inclusion in models of speech production.

Experiments

Our experiments extend the earlier observations by Lehiste (1973) and Coleman (1974) on listeners' inability to perceive the rhythm of speech veridically. We have used two tasks, a rhythm

matching task and a tapping task. The first two experiments used the rhythm matching task. Subjects adjusted the times between four noise bursts to match the overall rhythm of either a sentence or a control sequence of non-speech sounds. They could listen to the sound whose rhythm they were to match or the adjustable noise burst sequence by pressing one or other of two buttons; thus they could not hear the two stimuli simultaneously but were able to listen to each separately as many times as they liked while making the adjustments.

The sentence used in Experiment 1 was "A bird in the hand is worth two in the bush", synthesized on PAT from parameters derived from real speech. The noise burst sequence that subjects adjusted had four strong bursts corresponding to the four stressed syllables with appropriate intervening weaker bursts representing the unstressed syllables. By adjusting either of three knobs subjects could adjust the interval between adjacent stressed bursts, but the rhythm of the intervening weaker bursts was kept constant, scaled in tempo to the new inter-stress interval. Subjects matched the rhythm of two versions of the sentence; one had the natural pitch contour, the other a monotone. They also matched the rhythm of a control sequence of tones whose onsets were at the same time intervals as the stressed syllables of the sentence. Subjects performed seven matches (the first two of which were not analysed) to each of the three stimuli. The control was always done first followed by the two speech conditions in an order counter-balanced between subjects.

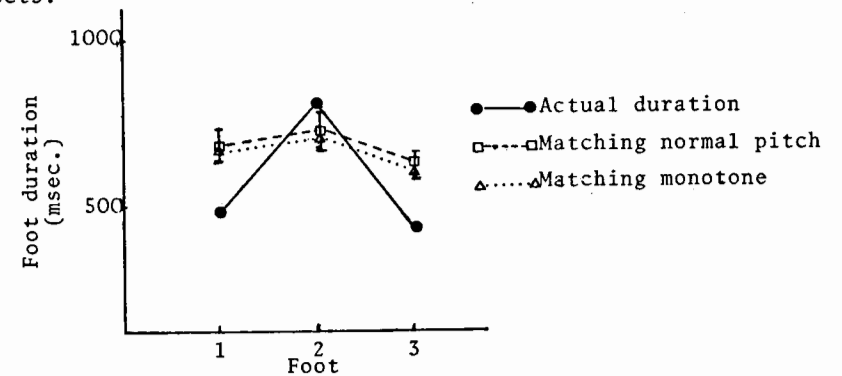


Fig.1. Actual and perceived foot durations for utterance in Experiment 1. (vertical bars represent  $\pm$  1 S.E. of the mean)



The actual and mean matched durations of the first three feet (that is the intervals between the four stressed syllables) are shown in Figure 1. To test any tendency for the matched durations to be more isochronous than the original utterance the quantity:  $|(1-a_i/a_{i+1})| - |(1-p_i/p_{i+1})|$  where  $a_i =$  actual duration of  $i^{\text{th}}$  foot,  $p_i =$  matched " " " " was calculated for  $i=1,2$ . A positive value for this quantity indicates that the perceived durations are more isochronous (over the two feet in question) than the actual durations. Such a tendency towards perceptual isochrony was reliable ( $p<.001$ ) for both foot-ratios when subjects matched the two sentences, but was not found when they matched the control, non-speech rhythm. The natural and the monotone speech are both perceived as more isochronous than they really are, but the non-speech tonal pattern is not.

The second experiment differed from the first as follows:

- 1) Four sentences of natural (female) speech were used which contained different numbers of syllables in each foot.
- 2) The stressed syllables in each utterance all began with a stop consonant (/t/) and there were no other occurrences of this sound in the utterance. This made it easier to specify to the subjects where the major stresses fell as, instead of saying match the rhythm of the 'syllable beats' or the 'tapping points', they could be told to hit the /t/'s.
- 3) The noise-burst sequence was made up of five bursts only; an initial low amplitude burst corresponding to the first, unstressed syllable in each utterance, and four 'stressed' bursts corresponding to the four stressed syllables.
- 4) Subjects were explicitly encouraged to use a strategy that we had observed in the first experiment, namely repeating the sentence to oneself while listening to the noise bursts. In case subjects' own articulations were more isochronous than the original, recordings were made of each subject speaking each sentence. In fact we found they were not more isochronous and the following results still hold if matched durations are compared with subjects' own productions.

For the four sentences as a whole, four of the eight foot-ratios gave a significant ( $p<.01$ ) tendency towards perceived isochrony, three gave no difference (partly because their foot ratios were actually quite close to unity already) and one gave a significant tendency away from perceptual isochrony. The results from

this deviant sentence and from one of the others are shown in Figure 2. Notice first in the right-hand panel that although subjects' judgements are quite reliable they are massively inaccurate at judging the duration of the middle foot. It is not clear though whether this huge overestimation of the middle foot should be attributed to perceptual isochrony. If it were, then we would not expect the similar, though more variable overestimation of the middle foot found in the left-hand panel for a sentence whose middle foot is already relatively long. Alternative explanations, which could account for the data from all four of the sentences, are that subjects overestimate the length of a foot containing (a) a major syntactic boundary or (b) a tone group boundary. Our third experiment looks at this question.

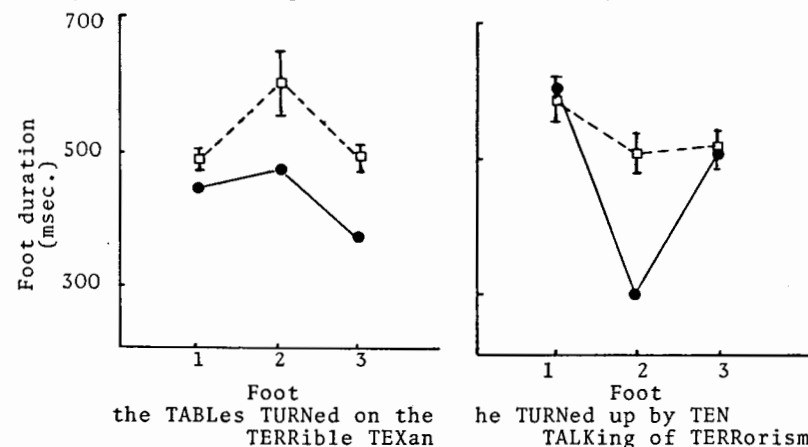


Fig.2. Actual and perceived foot durations for two of the utterances in Experiment 2.

To investigate the possible contribution of intonation to perceived rhythm a change of experimental technique was required. Subjects' imitations of sentences were, as we have seen, no more isochronous than the originals, but they did differ markedly in intonation. To ensure that subjects matched a sentence with the original intonation we asked them to tap in time to a sentence.

This new task differs from that used by Allen (1970) in that subjects tapped to every stressed syllable rather than just to a selected one on each trial. Here we are interested in subjects' perception of the entire rhythmic pattern. Subjects were not explicitly told to tap on the stressed syllables so the fact that they did provides an objective verification of the notion of the

rhythmic foot. The three utterances, which differed in number of tone groups and in syntactic structure, but which had identical foot durations were:

- 1) //1 Tim's in / Tuscany's / Training / Troops //
- 2) //1 Tim's in / Tuscany / Training / Troops //
- 3) //1 Tim's in / Tuscany //1 Training / Troops //

Here, following Halliday (1967), we bound tone groups by double slashes and indicated the type of tone group by a number. Both 2) and 3) contain a major syntactic boundary in the middle foot but in utterance 2) this was not marked by a tone group boundary. 1) and 2) were acoustically identical except that the /s/ of "Tuscany's" was spliced out for sentence 2) and replaced by four additional pitch periods of /v/ and an appropriate amount of silence to maintain the same foot length.

Fifteen subjects were divided into three groups, each group receiving a different order of presentation of the three utterances. Subjects heard each utterance 15 times and were told to start tapping after the third token. Each utterance was preceded by a warning tone 750 msec. from the onset of the utterance. Only the last 10 trials in each condition were analysed. Before each block of 15 trials, subjects heard the utterance three times and were given a context in which the utterances could occur. For example 3) might be the response to the question "Where's Tim and what's he doing?", while the same utterance with one tone group (sentence 2) might be the response to the question "What's Tim doing with the troops in Tuscany?" This was done to ensure that the subjects had a good idea of the syntax and tone group structure of the sentences they were listening to.

The results of this experiment (Figure 3) showed that while the number of tone groups has a distinct effect on perceived rhythm, the syntactic structure does not. In particular we found no tendency towards perceived isochrony in sentence 3, which contained two tone groups, but we did find a significant ( $p < .01$ ) tendency towards perceived isochrony for both foot-ratios in sentence 1 and 2. Sentences 1 and 2 did not differ from each other significantly in this respect, but both differed significantly from sentence 3 ( $p < .01$ ).

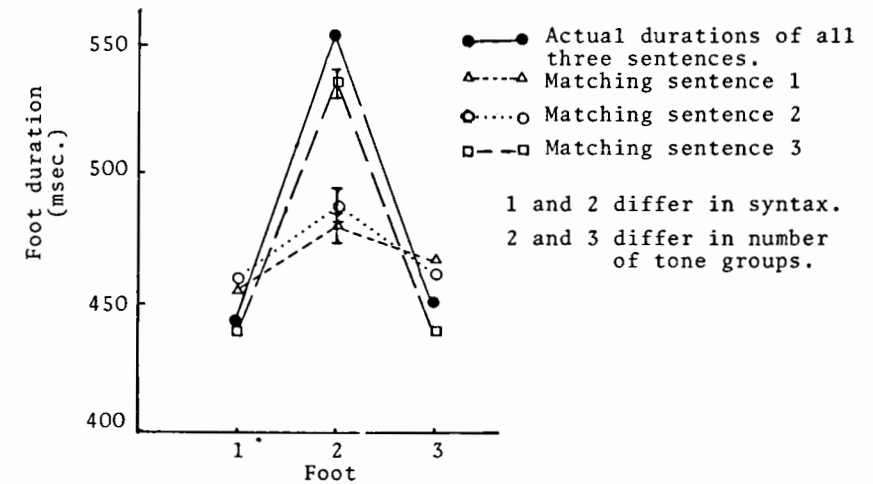


Fig. 3. Actual and perceived foot durations for the three utterances in Experiment 3 (see text for details).

It is apparent from these results that subjects are responding differently to the one and two tone-group utterances irrespective of the syntactic structure and despite the fact that the foot durations are the same in all three cases. Rees (1975), building on Halliday's (1967) work, has proposed that the tone group is a unit of rhythm as well as a unit of intonation so that isochrony need be maintained within but not between tone groups; it may be that this puts constraints on the limits of perceptual isochrony as well as on the tendency towards isochrony in production. It is clear from the experiments reported here that people are consistently inaccurate when judging speech rhythms and, furthermore, that they tend to hear these rhythms as more regular than they really are, at least when the utterance is bounded by a single tone group. Within the tone group, long feet tend to be underestimated, even when they contain a major syntactic boundary, while short feet tend to be overestimated.

#### Conclusions

Our results have broadly confirmed Lehiste's proposal that isochrony is partly a perceptual phenomenon. But we would make two points in addition. First, it is a perceptual phenomenon which is not independent of intonation. Second, we feel that it is a perceptual phenomenon, confined to language, reflecting underlying processes in speech production. Our results strengthen the case for models of the timing of English that incorporate an underlying

rhythmic organisation within tone groups. Conversely, they question the value of seeking direct links between syntax and segmental durations rather than indirect ones via an overall rhythmic structure which is also determined by the pragmatic and semantic context of a sentence (cf. Cutler & Isard, in press).

#### References

- Abercrombie, D. (1964): "Syllable quantity and enclitics in English", in In Honour of Daniel Jones, D. Abercrombie, D.B. Fry, P.A.D. MacCarthy, N.C. Scott, J.L.M. Trim (eds.) London: Longmans, 216-222.
- Allen, G. (1970): "The location of rhythmic stress-beats in English: An experimental study", UCLA Working Papers 14, 80-132.
- Allen, G. (1975): "Speech rhythm: its relation to performance universals and articulatory timing", JPh 3, 75-86.
- Coleman, C. (1974): A study of acoustical and perceptual attributes of isochrony, Ph.D. thesis, Univ. Washington.
- Cutler, A. and S.D. Isard (in press): "The production of prosody", in Language Production, B. Butterworth (ed.) New York: Academic Press.
- Fowler, C.A. (1977): Timing Control in Speech Production, Ph.D. thesis, Univ. Connecticut, Connecticut, Ind: Indiana Univ. Linguistics Club.
- Halliday, M.A.K. (1967): Intonation and Grammar in British English, The Hague: Mouton.
- Huggins, A.W.F. (1975): "On isochrony and syntax", in Auditory Analysis and Perception of Speech, G. Fant and M.A.A. Tatham (eds.), London: Academic Press, 455-464.
- Klatt, D. (1975): "Vowel lengthening is syntactically determined in a connected discourse", JPh 3, 129-140.
- Lehiste, I. (1973): "Rhythmic units and syntactic units in production and perception", JASA 54, 1228-1234.
- Lehiste, I. (1977): "Isochrony reconsidered", JPh 5, 253-263.
- Lindblom, B. and K. Rapp (1973): "Some temporal regularities of spoken Swedish", Papers from the Institute of Linguistics, University of Stockholm.
- Rees, M. (1975): "The Domain of Isochrony", Edinburgh Univ. Dept. of Linguistics, Work in Progress, 8, 14-28.
- Witten, I.H. (1977): "A flexible scheme for assigning timing and pitch to synthetic speech", L & S 20, 240-260.

## TEMPORAL ORGANIZATION OF SEGMENTAL FEATURES IN JAPANESE DISYLLABLES

Hiroya Fujisaki and Norio Higuchi, Faculty of Engineering,  
University of Tokyo, Tokyo, Japan

Introduction

While it is apparent that the realization of successive units in connected speech is based on the proper timing of articulatory and phonatory events, much remains to be investigated regarding the nature of the timing mechanism. It is not even agreed whether the regularity of timing (isochrony) resides in speech production or in speech perception, as pointed out by Lehiste (1977). The lack of our knowledge on this issue may primarily be due to the fact that the speech signal often fails to display marked segment boundaries, and that even the apparent boundaries do not directly reveal the timing of production nor the timing of perception. Elucidation of the mechanism underlying the isochrony thus requires experimental techniques for extracting, from the speech signal, the indices for the timing of production as well as the indices for the timing of perception of each of the successive units.

The present paper deals with both the productive and the perceptual aspects of the segmental timing in Japanese disyllabic words consisting only of vowels. Disyllabic words were selected since they display the characteristics of connected speech on the smallest scale. Vowel sequences were chosen since their acoustic characteristics can be most clearly defined in terms of formant frequencies, and the articulatory transition from the initial vowel to the second vowel can be traced in the trajectories of their formant frequencies.

The Speech Material

The speech material consisted of 20 disyllables, i. e., all the possible pairs of the five Japanese vowels (/i/, /e/, /a/, /o/, and /u/), pronounced with the "flat-type" word accent. Among these disyllables, nine were meaningful with the given accent type, four were meaningful when pronounced with a different accent type, and the rest were nonsense words. A randomized list of 100 words, containing five tokens each of the 20 disyllables, was read by a male speaker of the Tokyo dialect of Japanese. These disyllables were pronounced in isolation at an interval of three seconds. The speech signal was sampled at 10 kHz with an accuracy of 11 bits/sample and stored in the magnetic tape memory of a digital computer.

Analysis of Segmental Timing at the Level of Speech Production

An LPC analysis was made of all the utterances to extract the frequencies and bandwidths of 11 poles, from which the first three formant frequencies were selected on the basis of bandwidth and the continuity of the trajectories. These trajectories were then used to estimate the onset of the transition from the initial to the second vowel.

The estimation was based on the model of the coarticulation process in connected vowels previously proposed by Fujisaki et al. (1974, 1977). As shown in Fig. 1, the entire production process for connected vowels is represented by a hypothetical linear system which converts the stepwise target formant frequencies of each vowel into actual formant trajectories. An analysis of observed formant trajectories has indicated that a good approximation can be obtained by a critically-damped second-order linear system.

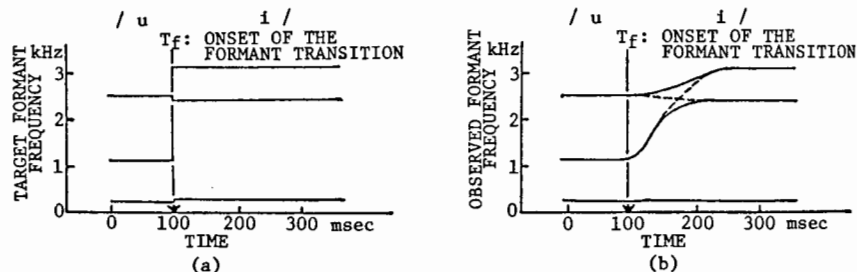


Fig. 1. Formulation of the process of coarticulation in the formant frequency domain: conversion of idealized formant target (a) into actual formant trajectories (b).

In the case of the disyllables under study here, we may assume a target frequency for the \$n\$th formant as

$$C_n(t) = F_{n1} + (F_{n2} - F_{n1}) u(t - T_f),$$

where \$F\_{ni}\$ denotes the target frequency of the \$n\$th formant of the \$i\$th vowel, and \$T\_f\$ denotes the onset of the transition measured from the voice onset of the initial vowel as the origin of the time axis. Then the actual formant frequency can be given by

$$F_n(t) = F_{n1} + (F_{n2} - F_{n1}) \{ 1 - (1 + \frac{t - T_f}{\tau_n}) \exp(-\frac{t - T_f}{\tau_n}) \} u(t - T_f),$$

where \$\tau\_n\$ denotes the time constant for the transition of the \$n\$th formant. Further considerations regarding the continuity and cou-

pling of the resonance modes lead to good approximations of the formant trajectories for all of the vowel combinations. When a set of observed formant trajectories (\$F\_1(t)\$, \$F\_2(t)\$, and \$F\_3(t)\$) is given, it is possible, by the method of Analysis-by-Synthesis, to determine the common onset of the formant transition and the time constants of individual formant trajectories. In the following analysis, a common time constant \$\tau\_2\$ was assumed for the second and third formants. Examples of the observed formant trajectories and their best approximations by the above-mentioned model are shown in Fig. 2 for /ui/ and /iu/, where the estimated onset \$T\_f\$ of the formant transition is also indicated.

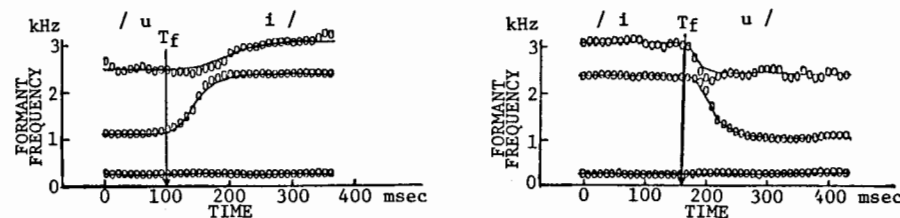


Fig. 2. Observed formant frequency trajectories (dots), their best approximations (—), and the estimated onset of the formant transition (\$T\_f\$) for /ui/ (left) and /iu/ (right).

Table 1 summarizes the results for all the utterance samples and lists the mean values of \$T\_f\$ and \$\tau\_2\$ for five tokens of each disyllable. The following comments can be drawn from a comparison of the results for pairs of disyllables having the same vowel combination in a different order.

first vowel	second vowel				
	/i/	/e/	/a/	/o/	/u/
/i/ \$T_f\$	-	155	131	136	134
\$\tau_2\$	-	21	25	22	27
/e/ \$T_f\$	90	-	132	134	149
\$\tau_2\$	59	-	40	26	20
/a/ \$T_f\$	119	125	-	124	125
\$\tau_2\$	48	39	-	41	37
/o/ \$T_f\$	101	94	92	-	113
\$\tau_2\$	44	58	51	-	-
/u/ \$T_f\$	108	117	126	146	-
\$\tau_2\$	39	44	28	-	-

Table 1. Mean values for the interval (\$T\_f\$[msec]) from voice onset to the onset of formant transition and for the time constant (\$\tau\_2\$[msec]) of the second formant trajectory for the 20 disyllabic words.

(1) In disyllables involving jaw movement without a change in lip articulation (i. e., /ie/ vs. /ei/, /ea/ vs. /ae/, /ia/ vs. /ai/, and /uo/ vs. /ou/),  $T_f$  is always larger for the disyllable produced by an opening movement of the jaw than for that produced by a closing movement. Analysis of variance indicates that the difference is highly significant (0.1% level) in /ie/ vs. /ei/, and is also significant (1% level) in /uo/ vs. /ou/, as well as in /ia/ vs. /ai/.

(2) In disyllables involving changes in lip articulation with or without minor jaw movement (i. e., /iu/ vs. /ui/, /eu/ vs. /ue/, /io/ vs. /oi/, /eo/ vs. /oe/, and /ao/ vs. /oa/),  $T_f$  is always larger for the disyllable produced by a rounding of the lips than for that produced by an unrounding of lips. The difference is significant in /eu/ vs. /ue/ (1% level); /ao/ vs. /oa/ (2% level); /eo/ vs. /oe/ (2% level); /io/ vs. /oi/ (5% level); and /iu/ vs. /ui/ (5% level).

(3) No significant difference in  $T_f$  was found for /au/ vs. /ua/, which involve both major jaw movement and changes in lip articulation in the transition from the initial to the second vowel. The effects of these two articulatory factors are considered to counteract and cancel each other.

These points may be easily observed in Fig. 3, which schematically shows the regions of the vowel target on the  $F_1 - F_2$  plane. An arrow from one vowel target to another corresponds to a disyllable,

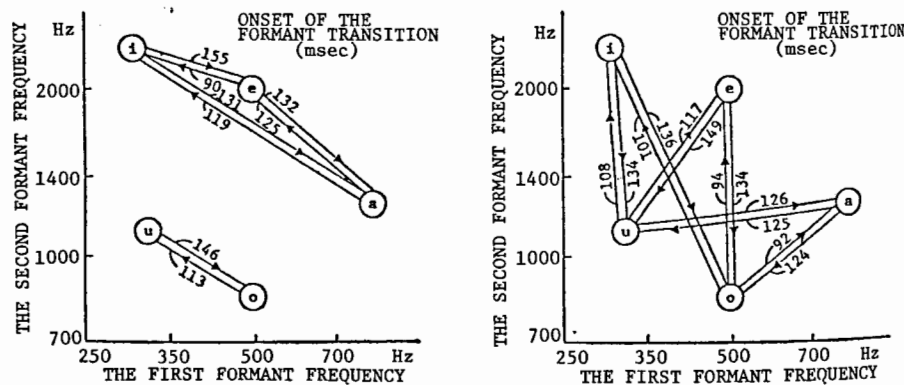


Fig. 3. Direction of the formant transition in the first and the second formant frequency plane and the onset of the formant transition ( $T_f$ ).

and the number associated with the arrow indicates the mean value of  $T_f$  (in msec) for that disyllable.

Furthermore, there exists a very high negative correlation between  $T_f$  and  $\tau_2$  ( $r = -0.91$ ) as shown in Fig. 4. Hence,

(4) Differences in the onset of transition ( $T_f$ ) tend to compensate for the differences in the rate of transition; a slower transition is initiated earlier and vice versa.

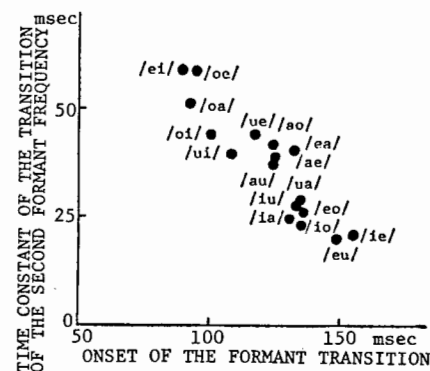


Fig. 4. Relationship between the onset of the formant transition ( $T_f$ ) and the time constant ( $\tau_2$ ) of the second formant trajectory.

#### Analysis of Segmental Timing at the Level of Speech Perception

The last finding of the preceding section suggests the possibility that the apparent diversity in the onset of transition in various disyllables is introduced to maintain the uniformity of the perceived duration of segments. The following experiment was designed to investigate this possibility, using the same utterance samples as in the above analysis to find the instant of the perceptual onset of the second vowel within a disyllable.

A set of 20 points were selected at intervals of 5 msec to cover the range of the major formant transitions in the waveform of each disyllabic utterance. Twenty tokens of truncated disyllables were then prepared by curtailing the original speech waveform at these 20 points. These tokens were arranged in serial order at an interval of 3.5 sec as stimuli in an identification test using the method of limits. The subject was asked to answer whether he heard one vowel segment or two in a truncated disyllable. The test was repeated to obtain the response probability, and the perceptual onset of the second vowel was defined as the point corresponding to an equal probability for the two alternatives. An example of the stimuli and the subject's response probability is schematically il-

illustrated in Fig. 5. The test was conducted using one utterance of each of the twenty disyllables. The subjects were two male speakers of the Tokyo dialect.

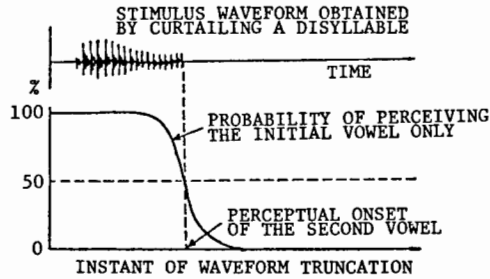


Fig. 5. Determination of the perceptual onset of the second vowel in a disyllable by waveform truncation.

Figure 6 shows the relationship between the perceptual onset ( $T_p$ ) of the second vowel and the onset of formant transition ( $T_f$ ) for each of the disyllables. Both  $T_p$  and  $T_f$  are expressed by their values relative to the total duration of an utterance. While the  $T_f$ 's for the various disyllables are distributed over a very wide range (22% - 42%), the  $T_p$ 's are found to be concentrated within a rather narrow range around the center of each utterance (48% - 53%). These findings suggest that the apparent diversity in the onset of the second vowel at the level of speech production may be the consequence of the speaker's effort to maintain the uniformity of perceived syllabic durations regardless of vowel combinations.

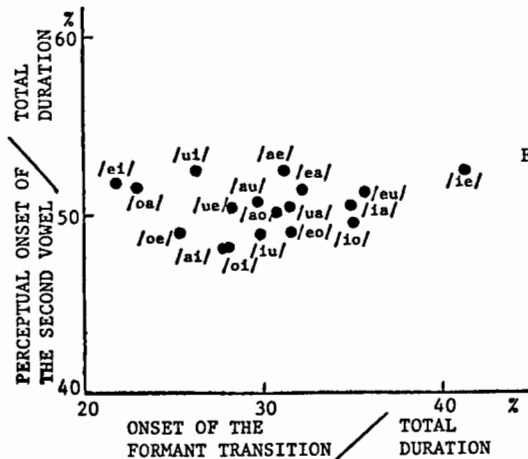


Fig. 6. Relationship between the perceptual onset ( $T_p$ ) of the second vowel and the onset of the formant transition ( $T_f$ ) for one sample of each of the disyllables.

Discussion

Two models of the possible mechanisms underlying the temporal organization of speech have been presented by Kozhevnikov and Chistovich (1965) and have since been widely discussed, e. g. by Ohala (1970), Leanderson and Lindblom (1972), and others. One is the so-called "chain model" based on the hypothesis of a closed-loop control of the speech production process. The other is the so-called "comb model" based on the hypothesis of an open-loop control. From our present knowledge concerning the motor organization of skilled behaviors, the chain model may be discarded, although it may certainly be true that various modes of feedback are necessary for the formation of the motor program. The findings of our present study suggest, however, that the comb model requires further elaboration. Our findings suggest that the formulation of the temporal relationship between the motor control and the articulatory/acoustic realizations of speech units is not complete without a consideration of their relationship to perceptual timing. From this point of view, two possible models can be distinguished under the open-loop (or "comb") hypothesis, as shown in Fig. 7.

In model (a), successive segments are produced with an isochronism at the level of the motor commands, so that their articulatory/acoustic realizations are not necessarily isochronous because of differences in the physiological and physical properties of the various articulators, as well as in the manner of articulation. In model (b), on the other hand, the motor commands and the articulatory/acoustic realizations of successive segments are programmed in such a way that the perceptual onsets of successive segments occur

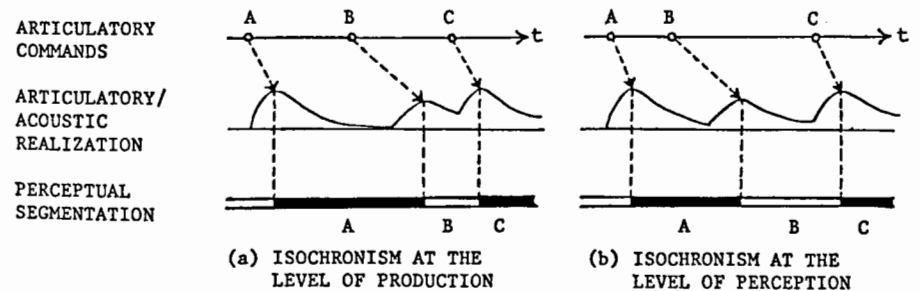


Fig. 7. Two models of the mechanisms underlying the temporal organization of speech units under the open-loop control hypothesis.

with an isochronism, viz., the perceived durations of these segments are kept equal. The results of the present study may be considered as corroborating model (b) as far as the Japanese disyllables are concerned.

#### Conclusion

Temporal organization of speech segments was investigated using disyllabic Japanese words consisting only of vowels. An acoustic analysis of their formant trajectories has indicated that the onset of the transition to the second vowel in various disyllables is distributed over a relatively wide range. This variation tends to compensate for the differences in the rate of transition due to differences in the articulator(s) involved and the direction of movement. On the other hand, a perceptual analysis of the onset of the second vowel has indicated that the perceptual onset of the second vowel in utterance samples of the same disyllable is concentrated within a relatively narrow range regardless of the particular vowel combination or the order of the vowels in the disyllable. The implication of these findings for the possible mechanisms underlying the temporal organization of speech units was discussed in connection with two models already proposed with regard to these mechanisms.

#### References

- Fujisaki, H. et al. (1974): "Formulation of the coarticulatory process in the formant frequency domain and its application to automatic recognition of connected vowels," Proc. SCS-74 3, 385-392.
- Fujisaki, H. (1977): "Functional models of articulatory and phonatory dynamics," in Articulatory Modeling and Phonetics, R. Carré, R. Descout, and M. Wajskop (eds.), 127-136, G. A. L. F. Group de la Communication Parlée.
- Kozhevnikov, V. A. and L. A. Chistovich (1965): Speech: Articulation and Perception, Moscow: Nauka.
- Leanderson, R. and B. E. F. Lindblom (1972): "Muscle activation for labial speech gestures," Acta Otolaryng. 73, 362-373.
- Lehiste, I. (1977): "Isochrony reconsidered," Journal of Phonetics 5, 253-263.
- Ohala, J. (1970): "Aspects of the control and production of speech," Working Papers in Phonetics 15, UCLA.



SOME EFFECTS ON INTELLIGIBILITY OF INAPPROPRIATE TEMPORAL  
RELATIONS WITHIN SPEECH UNITS

A. W. F. Huggins, Bolt Beranek and Newman Inc, 50 Moulton Street,  
Cambridge, Mass 02138, U. S. A.

The purpose of this paper is to make two arguments. The first is that, despite several failures to find such effects, badly disturbed speech timing, such as occurs often in the speech of the deaf for instance, is a sufficient cause for catastrophic loss of intelligibility. If the timing is sufficiently disturbed that the listener cannot identify the pattern of stressed syllables in the sentence -- or, perhaps, its rhythmic pattern -- the sentence will be unintelligible even though virtually all of the phonemes are clearly identifiable in subsequent listening. If the listener perceives a stress/rhythmic pattern that is different from that intended by the speaker, he is "garden-pathed" away from the correct utterance, and is not able to recode the individual phonemes into the words they represent before they fade from auditory short-term memory.

The second argument is that a reason for earlier failures to find strong relationships between timing and intelligibility is that a listener cannot estimate the effect of a particular timing distortion on speech intelligibility if he knows what the sentence says. This fact is already well known. It forms the basis of a popular way of impressing an audience with the fidelity of a speech vocoding system: a demonstration tape is prepared in such a way that the audience already knows what the test sentence is before they hear it as processed by the system whose performance is to be proved. What is not so well known is how easy it is to fall into the trap set by this fact. To be blunt, although I was very aware of the effect, I fell into the trap (Huggins, 1978), and if it can happen to me, it can happen to anyone!

Speech of the Deaf

A major reason for trying to understand speech timing is the need to improve the intelligibility of deaf speakers. Faulty timing has been implicated in poor intelligibility by virtually every major study of deaf speech this century, but this knowledge has not led to the development of effective training methods.

The most frequently cited ways in which the timing of deaf speech differs from normal speech are (1) slower overall rate; (2) more and longer pauses, often inappropriately placed; (3) inadequate differentiation of stressed and unstressed syllables; and (4) excessive lengthening of some segments, especially stops and fricatives (e.g. Nickerson, 1975). Let us consider the foregoing factors in order. Deaf speakers normally take much longer to produce a specified utterance than do normal-hearing speakers. But to the extent that the slower rate is a result of linear stretching of the time scale, slower speech should be more rather than less intelligible. One usually speaks slower (and also more precisely) to someone who has difficulty understanding, such as a child or a foreigner. Furthermore, when recorded speech is instrumentally expanded in time by a factor of four, intelligibility is not affected although the speech becomes tedious to listen to.

Similarly, it would be very surprising if the addition of appropriately placed pauses had a degrading effect on intelligibility. Pauses can be used to mark explicitly the boundaries between groups of syntactically related words. Boundaries so marked need not be inferred from more subtle cues, and the presence of syntactically appropriate pauses should therefore simplify rather than complicate reception. Further, the pauses effectively give the listener additional time to decode the message, and this too lightens rather than increases the processing load (Aaronson et al, 1971).

The occurrence of inappropriate pauses raises a different issue. Inappropriate pauses occur also in normal speech, where they are interpreted as hesitation pauses. These do not appear to interfere with intelligibility. However, listeners are much more sensitive to the presence of inappropriate than appropriate pauses, the threshold for their detection being almost five times smaller (Boomer and Dittmann, 1962). Presumably, then, if inappropriate pauses were interpreted as hesitation pauses in deaf speech also, no damage would result. Problems would arise, however, if the inappropriate pauses were interpreted as appropriate pauses, because this would signal incorrect segmentation of the message. This argument leads to rather a

different view of how timing errors might interfere with intelligibility: they might introduce misleading information about the message which, once accepted, could not be discarded.

There are other aspects of deaf speech which support such a view. Due to difficulties in coordinating different articulators, deaf speakers often produce sounds extraneous to the required sequence, particularly in making and releasing stops and fricatives (Hudgins and Numbers, 1942). If the listener accepts these extraneous sounds as segments, he cannot then go back and delete them. The perceptual apparatus is very good at filling in missing information, but it is very bad at discarding extraneous information unless it occurs as part of a separate auditory "stream" (Bregman and Campbell, 1971). Thus, listeners will swear that they heard a particular segment in a sentence even though it had been totally removed and replaced with an extraneous sound such as a cough (Warren et al, 1969). But the cough cannot be located in the sentence with any accuracy, since it cannot be integrated into a single stream with the speech. When wanted and unwanted segments arrive in a single auditory stream, as they often do in deaf speech, the listener cannot selectively accept the wanted and reject the extraneous segments, even if he had some way of so classifying the segments as they arrived. van Noorden (1975) has shown that two melodies in the same pitch range cannot be identified if they are played by interleaving the notes from the two melodies. The listener cannot decide to listen to alternate notes. On the contrary, he hears only a single sequence. But if one melody is gradually raised in pitch, the two melodies eventually split into two streams, permitting one to be ignored so that the other melody can be recognized.

The listener is not able to discard some of the information after it has been processed, either, and recent models of speech perception offer an explanation. Jarvella (1971) has shown that the accuracy of a listener's verbatim memory for a continuously presented message shows a sharp drop at the preceding clause boundary, as if the need to keep the raw acoustic data available in short-term memory ends when the clausal material is successfully parsed. Thus, any misinterpretations of the

preceding clause that become apparent later cannot easily be corrected, since the verbatim material necessary to the correction has been deleted from short term memory. Furthermore, if the received sequence of segments fails to trigger recognition of a word, the segments fade quite rapidly from auditory short term memory.

When the foregoing arguments are put together with the known importance of correct stress patterns for recognition of words, the poor intelligibility of deaf speech becomes much easier to understand. The pattern of stresses in a word or phrase is of critical importance to its correct recognition. In fact, there is evidence that listeners will discard correctly-heard segmental cues which they cannot reconcile with the perceived stress pattern. English listeners trying to identify English words and phrases, spoken with inappropriate stress patterns by Indian speakers, consistently produced words that matched the incorrect stress patterns, while correct phonemes occurred in enough of the responses to demonstrate that the necessary segmental cues were in fact present (Bansal, 1966). Second, it is known that timing is a vital cue in the perception of stress, outweighing both intensity (loudness) and pitch (Fry, 1958).

Yet it is not clear how much deaf speakers know about stress patterns. For normal listeners, the stress pattern of a word is centrally involved in its memory coding (Brown and McNeill, 1966). It is unlikely that the deaf use a similar coding without being explicitly taught it. Deaf children do not code letters, presented visually in an immediate recall task, in terms of their auditory and articulatory properties, as do normal hearing children and adults (Conrad and Rush, 1965). If the deaf subjects do not use an auditory or articulatory coding scheme for segments, it is very likely that they also use a different coding scheme for stress patterns -- if, indeed, they have a coding scheme for stress patterns at all. Unless the stress pattern of a word is a central part of its representation in memory, the stress pattern is not likely to be reflected in the required pattern of syllable timing when the word is spoken. Yet this pattern of syllable timing is crucial to the intelligibility of the word for hearing listeners.

There are two aspects of incorrect timing that should be distinguished. One type can be traced directly to the difficulty of programming a rapid sequence of articulations. Timing errors become more frequent and more severe as the sentence to be uttered is made more difficult to articulate. The remedy may lie in trying to teach words as integrated motor patterns, and practicing their production first in isolation and then by substituting them in overlearned phrase or sentence frames. This is particularly important in the case of function words, whose fluency in deaf speech is a major determinant of intelligibility (Monson and Leiter, 1975). Timing errors of the foregoing type could be labeled errors of performance, since the deaf speaker is presumably at least partly aware that his production has fallen short of what was intended. The other aspect of incorrect timing is more important, and errors of this type could be labeled errors of intention. Errors of intention occur if the deaf speaker's model of how speech should be timed is different from that of a hearing speaker. In particular, the model may not incorporate the rules for assigning relative stress levels, and for realizing these in timing patterns.

Some evidence supporting the importance for intelligibility of differentiating stressed and unstressed syllables has been reported by Osberger (1978). She produced slight improvements in intelligibility by editing deaf speech waveforms to correct inadequate differentiation of stressed and unstressed syllables. Her method, however, was unable to separate errors of performance from errors of intention, which may account for the smallness of her effects. Also, she reported no attempt to relate the magnitude of the timing corrections made in individual words to the resulting changes in intelligibility.

I have reported elsewhere a preliminary attempt to measure the effects of errors of intention uncontaminated by errors of performance, using synthetic speech (Huggins, 1978). Simple sentences were synthesized in two versions. In one, stress was correctly assigned, and in the other, unstressed syllables were assigned primary stress, and vice versa. Syllables with secondary stress were not affected. Since the same set of synthesis rules were used for stressed as for unstressed

syllables, any errors of performance that were inherent in the synthesis procedure should have affected the normal and mis-stressed versions equally. But when stress was wrongly assigned, word intelligibility fell from 85% to 50%, and the percentage of sentences "substantially understood" fell from 75% to 25%. The results were not uniform across test sentences, in part because the sentences differed in the proportion of syllables carrying primary, secondary, and un-stress, and in part because of some residual errors in phonetic transcription of the test sentences (which may well account for the less than perfect intelligibility of the normally stressed versions). I hope to correct some of these weaknesses in time for the meeting.

Finally, I want to repeat an anecdote from the study. I have tried several times to make a tape demonstrating how unintelligible speech can become when its timing is wrong, but I have never been satisfied with the results. In fact, I began to wonder if what I was trying to show was true. But when I played the latest tape to a colleague, looking for sympathy, he found it totally unintelligible. The difference between us was that I knew what each test sentence said, and therefore knew its stress pattern, whereas he did not. I would never have run the formal experiment but for his unexpected reaction. How many interesting timing effects have been overlooked, or regarded as too slight to be of interest, for similar reasons?

#### References

- Aaronson, D., N. Markowitz, and H. Shapiro (1971): "Perception and immediate recall of normal and "compressed" auditory sequences," Perception and Psychophysics, 9, 338-344.
- Bansal, R. K. (1966): The intelligibility of Indian English: measurements of the intelligibility of connected speech, and sentence and word material, presented to listeners of different nationalities, Unpublished Ph. D. Thesis, London University.
- Boomer, D. S. and A. T. Dittmann (1962): "Hesitation pauses and juncture pauses in speech," Language and Speech, 5, 215-220.
- Bregman, A. S. and J. Campbell (1971): "Primary auditory stream segregation and perception of order in rapid sequences of tones," J. Experimental Psychology, 89, 244-249.
- Brown, R. and D. McNeill (1966): "The tip of the tongue phenomenon," J. Verbal Learning and Verbal Behavior, 5, 325-337.

- Conrad, R. and M. L. Rush (1965): "Nature of short-term memory encoding by the deaf," J. Speech and Hearing Disorders, 30, 335-343.
- Fry, D. B. (1958): "Experiments on the perception of stress," Language and Speech, 1, 126-152.
- Hudgins, C. V. and F. C. Numbers (1942): "An investigation of intelligibility of speech of the deaf," General Psychology Monograph, 25, 289-392.
- Huggins, A. W. F. (1978): "Speech timing and intelligibility," in J. Requin (ed), Attention and Performance VII, Hillsdale, N.J.: Erlbaum.
- Jarvella, R. (1971): "Syntactic processing of connected speech," J. Verbal Learning and Verbal Behavior, 10, 409-416.
- Monson, R. B. and E. Leiter (1975): "Comparison of intelligibility with duration and pitch control in the speech of deaf children," J. Acoust. Soc. Amer., 57, S69 (A).
- Nickerson, R. S. (1975): "Characteristics of the speech of deaf persons," Volta Review, 77, 342-362.
- Osberger, M. J. (1978): The effect of timing errors on the intelligibility of deaf children's speech. Unpublished doctoral thesis, City University of New York.
- van Noorden, L. P. A. S. (1975): Temporal coherence in the perception of tone sequences. Eindhoven, Netherlands: Technische Hogeschool (Doctoral thesis).
- Warren, R. M., C. J. Obusek, R. M. Farmer, and R. P. Warren (1969): "Auditory sequence: confusions of patterns other than speech or music," Science, 164, 586-587.

## SYNTHESIS BY RULE OF SEGMENTAL DURATIONS IN ENGLISH SENTENCES

Dennis H. Klatt, Mass. Inst. of Tech., Cambridge, MA 02139.

In this paper, we are concerned with prediction of the (acoustically defined) durations of phonetic segments in spoken sentences. The durational definitions that have been adopted correspond to the closure for a stop (any burst and aspiration at release are assumed to be a part of the following segment). For fricatives, the duration corresponds to the interval of visible frication noise (or to changes in the voicing source if no frication is visible). For sonorant sequences, the segmental boundary is defined to be the half-way point in the formant transition for that formant having the greatest extent of transition. The definitions represent a convenient largely reproducible measurement procedure, but the physiological and perceptual validity of these boundaries have not been established.

In a review of the factors that influence segmental durations in spoken English sentences (Klatt, 1976a and references cited therein), it was concluded that only some of the systematic durational changes were large enough to be perceptually discriminable. The goal of this paper is to describe these first-order effects by rules.

Input Representation for a Sentence

The durational rule system to be presented is a part of a speech synthesis by rule program (Klatt, 1976b). The phonological component of this program accepts as input an abstract linguistic description of the utterance to be synthesized. The output of the phonological component is a detailed phonetic and prosodic representation of the utterance, including an acoustic duration for each segment. The symbol inventory is shown in Table 1; it includes 52 phonemes, 3 stress markers, 3 types of boundary indicators, and 6 syntactic structure indicators. An example of the use of some of these symbols is provided in Figure 1.

Phonemic Inventory. A traditional phonemic analysis of English is assumed, except that:

- (a) Vowel+/R/ syllables are transcribed with the special vowel nuclei /IR/ ("beer"), /ER/ ("bear"), /AR/ ("bar"), /OR/ ("boar"), and /UR/ ("pure"). Words like "player" and "buyer" should be transcribed with two syllables, i.e. /EY+/RR/ and /AY+/RR/.

(M #F DH AX			#C 1 OW L D	#C M 1 AE N )N	#C S 1 AE T
#F IH N			#F AX	#C R 1 AA K RR )	
Phone	Stress	Dur	Phone	Stress	Dur
SI	0	200	AE	1	165
DH	0	40	DX	0	20
IY	0	85	IH	0	65
OW	1	145	N	0	50
LX	0	65	AX	0	65
D	0	35	R	1	80
M	1	70	AA	1	140
AE	1	225	K	0	50
N	0	60	RR	0	175
S	1	105	SI	0	200

Figure 1. Input representation for "The old man sat in a rocker" and a listing of the output of the phonological component, i.e. the phonetic string, stress feature, and duration predictions in msec.

- (b) The glottal stop [Q], dental flap [DX], glottalized alveolar stop [TQ], and velarized lateral [LX] listed in Table 1 are not really phonemes, but are allophones that are inserted in lexical forms before segmental durations are computed.

Lexical Stress. Each stressed vowel of an utterance must be preceded by a stress symbol (1, 2, or !), where 1 is primary lexical stress (reserved for vowels in open-class content words, only one 1-stress per word). The secondary lexical stress "2" is used in some content words (e.g. the first syllable of "demonstration"), in compounds (e.g. the second syllable of "baseball"), in the strongest syllable of polysyllabic function words (e.g. "until"), and for pronouns (excluding personal pronouns like "his"). Emphatic stress "!" can be assigned to a semantically prominent syllable in a phrase.

Morpheme and Word Boundaries. There is no input symbol to indicate a syllable boundary. The symbol "\*" can be used to mark morpheme boundaries. Each word of an utterance to be synthesized must be immediately preceded by a word boundary symbol. The distinction between content and function words is indicated by using "#C" and "#F". Open-class words (nouns, verbs, adjectives and adverbs) are content words. The program will check to see that no function word carries primary stress. A compound such as "apple cart" is indicated in the input representation by replacing the word boundary between "apple" and "cart" by a morpheme boundary and by reducing the lexical stress on the second word "cart" by one.

Table 1. The legal input symbols for synthesis of an utterance. Also given are a basic or inherent duration for each phonetic segment type and a minimum stressed duration in msec.

Vowels		INH DUR	MINDUR			INH DUR	MINDUR
IY	beet	160	50	IH	bit	130	40
EY	ba <u>i</u> t	190	70	EH	b <u>e</u> t	150	60
OW	bo <u>a</u> t	220	70	AH	b <u>u</u> t	140	50
UW	bo <u>o</u> t	210	60	UH	bo <u>o</u> k	160	50
AE	b <u>a</u> t	230	60	AA	Bo <u>b</u>	240	80
AO	bo <u>u</u> ght	240	80	RR	bi <u>r</u> d	180	60
AY	bi <u>t</u> e	250	90	AW	bo <u>u</u> t	260	100
OY	bo <u>y</u>	280	110	YU	be <u>a</u> uty	230	100
AX	ab <u>o</u> ut	120	40	IR	be <u>e</u> r	230	100
ER	b <u>e</u> ar	270	100	AR	ba <u>r</u>	260	100
OR	bo <u>a</u> r	240	100	UR	po <u>o</u> r	230	100
<u>Sonorant Consonants</u>							
W	w <u>e</u> t	80	60	Y	y <u>e</u> t	80	40
R	r <u>e</u> nt	80	30	L	l <u>e</u> t	80	40
WH	w <u>h</u> ich	70	60	H	h <u>a</u> t	80	20
EL	b <u>o</u> ttle	160	110	LX	b <u>i</u> ll	90	70
<u>Nasals</u>							
M	m <u>e</u> t	70	60	N	n <u>e</u> t	65	35
NG	s <u>i</u> ng	80	50	EM	k <u>e</u> e'p <u>e</u> m	170	110
EN	b <u>u</u> tten	170	100				
<u>Fricatives</u>							
F	f <u>i</u> n	120	60	V	v <u>a</u> t	60	40
TH	t <u>h</u> in	110	40	DH	t <u>h</u> at	50	30
S	s <u>a</u> t	125	50	Z	z <u>o</u> o	75	40
SH	sh <u>i</u> n	125	50	ZH	az <u>u</u> re	70	40
<u>Plosives</u>							
P	p <u>e</u> t	85	50	B	b <u>e</u> t	80	50
T	t <u>e</u> n	65	40	D	d <u>e</u> bt	65	40
K	c <u>o</u> re	65	50	G	g <u>o</u> re	65	50
DX	b <u>u</u> tter	20	20	TQ	at Alan	65	50
Q	Ma <u>o</u> pted	20	20				
<u>Affricates (closure, frication)</u>							
CH	ch <u>i</u> n	70	50	J	g <u>i</u> n	70	50
		60	40			30	20
<u>Stress Symbols</u>							
1	primary lexical stress						
2	secondary lexical stress						
!	emphatic stress						
<u>Word and Morpheme Boundaries</u>							
*	morpheme boundary						
#C	begin content word						
#F	begin function word						
<u>Syntactic Structure</u>							
.	end of declarative utterance						
)?	end of yes/no question						
(M	begin main clause						
,	orthographic comma						
)N	end of noun phrase						
(R	begin relative clause						

Syntactic structure. Syntactic structure symbols are important determiners of sentence stress, rhythm, and intonation. Syntactic structure symbols appear just before the word boundary symbol. Only one syntactic marker can appear at a given sentence position. The strongest syntactic boundary symbol is always used (the stronger symbols appear higher in the list in Table 1).

An utterance must end with either a period "." signalling a final fall in intonation, or a question mark ")" signalling the intonation pattern appropriate for yes-no questions. Each clause must be preceded by either "(M" to indicate the beginning of a main clause, or "(R" to indicate the beginning of a relative clause. If clauses are conjoined, a syntactic symbol is placed just before the conjunction. If a comma could be placed in the orthographic rendition of the desired utterance, then the syntactic comma symbol "," should be inserted. Syntactic commas are treated as full clause boundaries in the rules; they are used to break up larger units into chunks in order to facilitate perceptual processing. The end of a noun phrase is indicated by ")N". Segments in the syllable prior to a syntactic boundary are lengthened. Based on the results of Carlson, Granstrom, and Klatt (1979), an exception is suggested in that any )N following a noun phrase that contains only one primary-stressed content word should be erased. The NP + VP is then spoken as a single phonological phrase with no internal phrase-final lengthening.

#### Rules

The representation for a sentence discussed above serves as input to the phonological component of the synthesis-by-rule program. The form of the output from the phonological rules is shown at the bottom in Figure 1. The abstract string of symbols has been converted to a string of phonetic segments, with each segment being assigned a stress feature and duration in msec. Before presenting details of the duration algorithm, we summarize some of the rules that must be executed prior to duration prediction.

Stress Rules. The phonological component assigns a feature Stress (value = 0 or 1) to each phonetic segment in the output string. The default value is 0 (unstressed). Vowels preceded by a 1 or 2-stress in the input are assigned a value of 1. Consonants

preceding a stressed vowel are also assigned a value of 1 if they are in the same morpheme and if they form an acceptable word-initial consonant cluster. Segmental stress is used in rules that determine segmental duration, fundamental frequency, plosive aspiration duration, and formant target undershoot.

Rules of Segmental Phonology. There are presently very few phonological rules of a segmental nature in the program. A number of rules that are sometimes attributed by linguists to the phonological component (e.g. palatalization) are realized in the phonetic component because they involve graded phenomena (e.g. the [S] of "fish soup" is partially palatalized, but not identical to [SH:]. The segmental (within-word and across-word-boundary) phonological rules that are described below are extremely important. They are not "sloppy speech" rules, but rather rules that aid the listener in hypothesizing the locations of word and phrase boundaries. For example, the second rule ensures that a word-final /T/ is not perceived as a part of the next word by inserting simultaneous glottalization to attenuate any release burst. Rules are expressed in a feature-based notation that is compiled into Fortran code for computer simulation of the phonological component (Klatt, 1976b). Rules 1 and 2 below are stated in this way, while the others are expressed in ordinary English.

1. [L] --> [LX]/(+VOWEL)...(-STRESS)  
Substitute a postvocalic velarized allophone [LX] for [L] if the [L] is preceded by a vowel and followed by anything except a stressed vowel in the same word.
2. ([T] or [D]) --> [DX]/(+SONOR -NASAL)...(-STRESS +VOWEL)  
Replace [T] or [D] by the alveolar flap [DX] within words and across words boundaries (but not across phrase and clause boundaries) if the plosive is followed by a non-primary-stressed vowel and preceded by a nonnasal sonorant. Examples: "butter", "ladder", "sat about".
3. A word-final [T] preceded by a sonorant is replaced by the glottalized dental stop TQ (i.e. has a glottal release rather than a t-burst) if the next word starts with a stressed sonorant (unless there is a clause boundary between the words, in which case the [T] is released into a pause). Examples: "that one", "Mat ran".
4. A voiceless plosive is not released if the next phonetic segment is another voiceless plosive within the same clause.
5. A glottal stop [Q] is inserted before a word-initial stressed vowel if the preceding segment is syllabic (and not a determiner), or if the preceding segment is a voiced nonplosive and there is an intervening phrase boundary. Example: "Liz eats".

6. Unstressed [OR] is replaced by syllabic [RR], as in "for him" or "forget". (There are many rules of this type.)

Duration Rules. Each segment is assigned a duration by a set of rules presented in detail below. The rules are intended to match observed durations for a single speaker (DHK) reading paragraph-length materials. The rules operate within the framework of a model of durational behavior which states that (1) each rule tries to effect a percentage increase or decrease in the duration of the segment, but (2) segments cannot be compressed shorter than a certain minimum duration (Klatt, 1976a). The model is summarized by the formula:

$$DUR = ((INH DUR - MINDUR) * PRCNT) / 100 + MINDUR \quad (1)$$

where INHDUR is the inherent duration of a segment in msec, MINDUR is the minimum duration of a segment if stressed, and PRCNT is the percentage shortening determined by applying rules 1 to 10 below. The program begins by obtaining values for INHDUR and MINDUR for the current segment from Table 1, and by setting PRCNT to 100. The inherent duration has no special status other than as a starting point for rule application; it is roughly the duration to be expected in nonsense CVCs spoken in the carrier phrase "Say bVb again" or "Say CaC again". The following ten rules are then applied, where each rule modifies the PRCNT value obtained from the previous applicable rules according to the equation:

$$PRCNT = (PRCNT * PRCNT1) / 100 \quad (2)$$

The duration of the segment is then computed by inserting the final value for PRCNT into Equation 1 and, finally, Rule 11 is applied.

1. PAUSE INSERTION RULE: Insert a 200 msec pause before each sentence-internal main clause and at boundaries delimited by a syntactic comma, but not before relative clauses. The "(R" symbol functions like a "N" in the duration rules.
2. CLAUSE-FINAL LENGTHENING: The vowel or syllabic consonant in the syllable just before a pause is lengthened by PRCNT1=140. Any consonants between this vowel and the pause are also lengthened by PRCNT1=140.
3. NON-PHRASE-FINAL SHORTENING: Syllabic segments (vowels and syllabic consonants) are shortened by PRCNT1=60 if not in a phrase-final syllable. A phrase-final postvocalic liquid or nasal is lengthened by PRCNT1=140.
4. NON-WORD-FINAL SHORTENING: Syllabic segments are shortened by PRCNT1=85 if not in a word-final syllable.
5. POLYSYLLABIC SHORTENING: Syllabic segments in a polysyllabic word are shortened by PRCNT1=80.



6. NON-INITIAL-CONSONANT SHORTENING: Consonants in non-word-initial position are shortened by  $PRCNT1=85$ .
7. UNSTRESSED SHORTENING: Unstressed segments are half again more compressible than stressed segments (i.e. set  $MINDUR=MINDUR/2$ ). Then both unstressed and 2-stressed segments are shortened by a factor  $PRCNT1$  that is tabulated below for each type of segment. The result is that segments assigned secondary stress are shortened relative to 1-stress, but not as much as unstressed segments.

Context	PRCNT1 for Unstr. and 2-stress
syllabic (word-medial syll)	50
syllabic (others)	70
prevocalic liquid or glide	10
all others	70

8. LENGTHENING FOR EMPHASIS: An emphasized vowel is lengthened by  $PRCNT1=140$  percent.
9. POSTVOCALIC CONTEXT OF VOWELS: The influence of a postvocalic consonant or sonorant-stop cluster on the duration of a vowel is given below. (Cs must be in the same morpheme as the V and must have the feature unstressed.) In a postvocalic sonorant-obstruent cluster, the obstruent determines the effect on the vowel and on the sonorant.

Context	PRCNT1
open syllable, word-final	120
before a voiced fricative	160
before a voiced plosive	120
before an unstr. nasal	85
before a voiceless plosive	70
all others	100

The effects are greatest at phrase and clause boundaries: if non-phrase-final, change  $PRCNT1$  to be  $70 + 0.3*PRCNT1$

10. SHORTENING IN CLUSTERS: Segments are shortened in consonant-consonant sequences (disregarding word boundaries, but not across phrase boundaries), and segments are also modified in duration in vowel-vowel sequences.

Context	PRCNT1
vowel followed by a vowel	120
vowel preceded by a vowel	70
consonant surrounded by consonants	50
consonant preceded by a consonant	70
consonant followed by a consonant	70

11. LENGTHENING DUE TO PLOSIVE ASPIRATION: A 1-stressed or 2-stressed vowel or sonorant preceded by an aspirated plosive is lengthened by 25 msec.

When the rules are applied to the /RR/ of "rocker" in Figure 1, the second rule sets  $PRCNT$  to 140, the fifth rule reduces  $PRCNT$  to 112, the seventh rule reduces  $MINDUR$  to 30 msec and  $PRCNT$  to 78.4, and the ninth rule increases  $PRCNT$  to 94. Then  $INHUR$ ,  $MINDUR$ , and  $PRCNT$  are inserted in Equation 1 and the resulting duration is rounded up to the nearest 5 msec to obtain the value of 175 msec.

The resulting durations are determined in part by a variable that controls the nominal speaking rate  $SPRATE$  which can be set to any number between 60 and 300 words per minute. The default value is 180 words per minute. At rates slower than 150 wpm, a short pause is inserted between a content word and a following function word. (At a normal speaking rate, brief pauses are inserted only at the ends of clauses.) Individual segments are lengthened or shortened slightly depending on speaking rate, but most of the rate change is realized by manipulating pause durations.

#### Evaluation

The rules constitute only a first-order approximation to many of the durational phenomena seen in sentences (e.g. consonant interactions in clusters) and the rules completely ignore other factors. Nevertheless, as a first approximation, the rules capture a good deal of the systematic variation in segmental durations for speaker DHK. When compared with spectrograms of new paragraphs read by this speaker, the rule system produces segmental durations that differ from measured durations by a standard deviation of 17 msec (excluding the prediction of pause durations), and the rules account for 84 percent of the observed total variance in segmental durations. Seventeen msec is generally less than the just-noticeable difference for a single change to segmental duration in sentence materials (Klatt, 1976a).

A perceptual evaluation of the performance of the rule system is discussed by Carlson, Granstrom and Klatt (1979). The perceptual results are encouraging in that both naturalness and intelligibility ratings of sentences synthesized by these rules are very similar to ratings of the same sentences synthesized using durations obtained from a natural recording.

#### References

- Carlson, R., Granstrom, B., and Klatt, D.H. (1979), "Some Notes on the Perception of Temporal Patterns in Speech", 9th International Congress of Phonetic Sciences, Copenhagen.
- Klatt, D.H. (1976a), "Linguistic Uses of Segmental Duration in English: Acoustic and Perceptual Evidence", J. Acoust. Soc. Am. 59, 1208-1221.
- Klatt, D.H. (1976b), "Structure of a Phonological Rule Component for a Speech Synthesis by Rule Program", IEEE Trans. Acoustics, Speech, and Signal Processing ASSP-24, 391-398.



## COMPLEX CONTROL OF SIMPLE DECISIONS IN THE PERCEPTION OF VOWEL LENGTH

Sieb G. Nootboom, Institute for Perception Research, Eindhoven Netherlands

Introduction

Measurable vowel durations in connected speech are influenced by many different factors. Well-known examples are the effects on the durations of stressed vowels of (1) the postvocalic consonants, (2) the position of the syllable in the word, (3) the syntactic position of the word in the sentence, (4) the presence or absence of a speech pause following the syllable to which the vowel belongs, (5) overall speech rate. Despite the wide variability in acoustic durations due to the combined effects of these and other factors, in many languages vowel durations are contrastive cues to vowel phoneme perception. The fact that the shortest durations of phonemically long vowels can be considerably shorter than the longest durations of phonemically short vowels does apparently not prevent listeners from making correct decisions as to perceived phonemic vowel length. Let us examine how a decision strategy might be organized in order to accomplish this.

A hypothesized decision strategy

We assume, in terms of signal detection theory, that each acoustic vowel duration is represented on an internal stimulus strength axis with an uncertainty that is equal for all vowel durations, at least within the limited range of durations in which we are interested. This uncertainty, due to the effects of sensation noise, gives rise to a Gaussian distribution of stimulus strength values for repetitions of each particular vowel duration. We further assume that identification of vowel length is so organized that each individual stimulus strength value  $X$ , derived from a single vowel duration, is compared to an internal criterion  $C$  on the stimulus strength axis. All  $X$  lower than  $C$  are identified as short vowel phonemes, all  $X$  higher than  $C$  are identified as long vowel phonemes. The criterion  $C$  is assumed here to be noiseless so that all uncertainty in phonemic decisions has to be caused by the effects of sensation noise. Due to the Gaussian form of the stimulus strength distributions, the probability distribution of short vowel decisions over a set of acoustic vowel durations, and of course the complementary probability distribution of long vowel decisions,

will have the form of a cumulative normal distribution with a mean  $\mu$  (the phoneme boundary) reflecting the position of the internal criterion  $C$  on the stimulus strength axis, and a standard deviation  $\sigma$  reflecting the effect of sensation noise.

Although the internal criterion  $C$  is supposed to be noiseless, this does not imply that it has always the same position on the stimulus strength axis. We assume that the listener can move the internal criterion up and down the stimulus strength axis, adjusting its position in order to optimize the chance of correct recognition. Imagine that a listener perceives a vowel segment and is not sure whether the phoneme intended by the speaker was a long or a short vowel. He may then take into account that the vowel concerned is clearly stressed, is followed by a voiceless plosive, is in the final syllable of a word which is immediately followed by a major syntactic boundary, not accompanied by a speech pause, and that the speech rate is slightly faster than normal. Our listener knows from experience that in these conditions a short vowel would have had a stimulus strength value  $A$  and a long vowel a stimulus strength value  $B$ . He therefore places his internal criterion  $C$  in the middle between  $A$  and  $B$ , in this way optimizing his chance of a correct decision on perceived vowel length.

Of course, such a decision strategy implies that listeners have an extensive and detailed knowledge of the temporal regularities of speech and are able to apply this knowledge very rapidly, so rapidly in fact that they are not aware of doing so. They are even unaware of having this knowledge. We cannot, therefore, test the proposed theory by asking listeners what they do. We need another kind of test. Let us examine a few specific hypotheses that may be derived from the theory and see whether they are corroborated by experimental data.

The effect of sensation noise

If our theory is correct we could measure the accuracy of auditory representation in a vowel length identification task, by determining the  $\sigma$  of the cumulative normal distribution. Of course, in psychoacoustics the accuracy of auditory representation is generally expressed in terms of a differential threshold measured in a binary forced choice comparison task, involving two stimuli per decision. Our theory predicts that the accuracy is equal in both tasks, because there is no reason to suppose that the effect of

sensation noise would be different. In testing this hypothesis, however, we must take into account that in a binary forced choice task involving two stimuli the stimulus separation needed to obtain a given level of performance, for example the 75 % level, is  $\sqrt{2}$  greater than the stimulus separation needed to obtain the same level of performance in a similar task involving only one stimulus per decision (Green and Swets, 1966, 68). Because the  $\sigma$  of a cumulative normal distribution is almost  $\sqrt{2}$  times the 75 % differential threshold, we may compare the  $\sigma$  measured in a binary forced choice identification task directly to 75 % differential thresholds measured in a comparison task.

In a number of identification tests on the effect of acoustic vowel duration on the distinction between Dutch short /a/ and long /a:/ in different speech contexts, ranging from isolated vowels to vowels embedded in full sentences, we have found  $\sigma$ 's averaged over groups of at least 10 listeners for each context condition, between 10 ms (for vowels in isolation) and 3 ms (for one particular full-sentence condition). In most conditions  $\sigma$ 's were in the order of 6 ms. Phoneme boundaries ranged from 75 to 100 ms. These  $\sigma$ 's are within the range of differential thresholds of sound duration reported in the literature for sounds with approximately the same durations (Lehiste, 1970; Nootboom and Doodeman, 1978). There is an unpredicted and clear tendency for the  $\sigma$ 's to decrease when more speech context becomes available to the listeners. If we stick to our assumption that, within each particular context, the internal criterion C is noiseless, this would mean that the effect of sensation noise is context dependent. If one would prefer to assume that the effect of sensation noise is independent of speech context, one would have to abandon the idea of a noiseless criterion, and assume that the internal criterion shows less uncertainty with increasing embeddedness of the vowel segment.

#### Moving the internal criterion up and down

Let us now see what happens to the phoneme boundary, being the measurable reflection of the assumed internal criterion C in the decision strategy, when we change the speech context. The proposed strategy implies that, when the speech context changes in such a way that the expected durations of short and long vowel phonemes change, the internal criterion will move up and down accordingly. This prediction has been tested for changes in speech context re-

lated to (1) the postvocalic consonant, (2) the position of the syllable in the word, (3) the syntactic position of the word in the sentence, (4) the presence or absence of a speech pause after the syllable, (5) the overall speech rate. The experimental design was very similar in all cases and has been described in detail elsewhere (Nootboom and Doodeman, 1978). All experiments were limited to the Dutch /a/ - /a:/ opposition. It should be noted that in natural speech these two vowels are distinguished not only by their relative durations but also by their spectral properties, which were kept constant in the experiments. All experiments involved at least 10 subjects. Let us briefly review the results.

#### - The postvocalic consonant

Vowel segments followed by a speech pause have generally a greater duration than those followed by a consonant. The amount of shortening caused by the postvocalic consonant depends on the nature of the consonant. For example, plosives tend to shorten the preceding vowel more than do fricatives. This is valid for both short and long vowel phonemes. Consequently the optimal position of the internal criterion C will be at a lower stimulus strength value for vowels followed by a fricative than for vowels in isolation, and at a still lower value for vowels followed by a voiceless plosive. We therefore can predict that the phoneme boundary between /a/ and /a:/ in isolation lies at a greater duration than the same phoneme boundary measured before /s/, which again lies at a greater duration than the one measured before /t/. This prediction is corroborated by the data. We find the following phoneme boundaries, estimated from probability distributions in a vowel length identification test by fitting cumulative normal distributions (sd stands for the standard deviation over the subjects, and is not to be confused with the earlier discussed  $\sigma$ ):

In isolation 100 ms (sd 8.4 ms)

Before /s/ 97 ms (sd 6.7 ms)

Before /t/ 91 ms (sd 6.9 ms)

#### - The position in the word

Both short and long Dutch vowels bearing lexical stress tend to become shorter with increasing number of unstressed syllables following in the word (Nootboom, 1973). Thus we predict that the phoneme boundary will shift towards shorter durations when more

unstressed syllables are added to the word. This has been tested with nonsense words in isolation, in which the first, stressed, syllable contained the test vowel segment, followed by intervocalic /t/, and the number of unstressed syllables was 0, 1, 2 or 3. Phoneme boundaries and standard deviations over the subjects were:

- 0 unstr. syll. 91 ms (sd 6.9 ms)
- 1 unstr. syll. 88 ms (sd 5.8 ms)
- 2 unstr. syll. 85 ms (sd 5.5 ms)
- 3 unstr. syll. 83 ms (sd 4.3 ms)

Differences between the last three conditions might have been induced by perceived changes in the second syllable.

- The syntactic position of the word

Durations of Dutch short and long vowels in monosyllabic words vary systematically with syntactic position of the word, notably with the type of syntactic boundary immediately following the word. In one experiment we measured phoneme boundaries between /a/ and /a:/ in a monosyllable /tVk/ embedded in 5 different test utterances, each with a different syntactic structure. These test utterances were obtained from sentences spoken with the long vowel /a:/ in the test segment slot, and had normal rhythm and intonation. This original vowel /a:/ had durations ranging from 150 ms in one utterance to 190 ms in another. In each spoken sentence the original vowel segment was excised and replaced by one of a set of test segments differing in acoustic duration. Phoneme boundaries were assessed in each of the 5 test utterances for each of 12 subjects. They ranged from 76 ms, for the test utterance with an original /a:/- duration of 150 ms, to 100 ms, for the test utterance with an original /a:/-duration of 190 ms. Calculating the product moment correlation of the original /a:/-durations in each of the 5 test utterances with all 12 phoneme boundaries in each of these utterances gave  $r = 0.83$  ( $p < 0.001$ ). Apparently the /a:/-durations as originally spoken in these test utterances are to a fair extent controlled by the same factors as the phoneme boundaries in the identification test. These factors are either to be looked for in the syntactic structures of the sentences, as intended by the speaker and perceived by the listeners, or, more probably, in the prosodic structures of the sentence realizations, which, of course, are partly determined by the syntactic structures.

- The prepausal position of the syllable

Durations of Dutch short and long vowels in monosyllabic words are considerably longer when the word is in prepausal position than when it is not, other things being equal. Thus we may predict that the phoneme boundary in an embedded monosyllable will shift towards a greater duration when we insert a speech pause immediately after the monosyllable. This has been tested by inserting acoustic silent intervals with durations of 0, 100, 200, and 800 ms immediately after the test syllable /tVk/ embedded in a test utterance, and measuring the phoneme boundary in each of these conditions. In addition, the probability of speech pause perception has been measured independently. It was found that the phoneme boundary increased from 79 ms for a silent interval of 0 ms, to 94 ms for a silent interval of 800 ms, and that the phoneme boundaries in all test conditions were accurately predicted by

$$pb = 79 + 15P_{spp}$$

in which pb is the phoneme boundary in ms, and  $P_{spp}$  is the probability of speech pause perception. The probability distribution of speech pause perception over the durations of silent intervals follows an exponential function with a time constant of 200 ms. One possible interpretation of these results is that the listeners employed in this experiment two discrete internal criteria for vowel length identification, one for the prepausal and one for the non-prepausal condition. The gradual increase of the mean phoneme boundary value could be entirely due to the gradual increase in the probability of speech pause perception.

- The effect of speech rate

The effect of speech rate was assessed in a listening experiment employing a computer-controlled channel vocoder in order to vary overall speech rate of a test utterance. In this test utterance the syllable /tVk/ was in final position and the speech rate of all speech material preceding the test vowel segment was 0.67, 1, or 1.5 times normal. Phoneme boundaries were 81, 95 and 102 ms respectively, showing a partial adjustment to speech rate.

Conclusions

The proposed decision strategy for the disambiguation of vowel length is confirmed in all experimental tests. The effect of sensation noise does not conflict with what would be predicted from

differential thresholds for sound duration, although there is an unexpected tendency towards higher accuracy with increasing amount of embeddedness of the test vowel segment. The effect of speech context on the phoneme boundaries is found to be in the predicted direction in all 5 types of contextual differences which were investigated.

The shifts of the phoneme boundary may seem small, ranging from a few ms to about 25 ms. However, they are generally in the same order of magnitude as contextual effects on the durations of short vowels. Analogous to the tendency towards higher within-subject accuracy with increasing amount of embeddedness, we find less inter-subject variation with increasing amount of embeddedness, suggesting that the position of the internal criterion C becomes more and more constrained by inter-personal factors when more speech context becomes available to the listener.

The results support our hypothesis that listeners, whether they know it or not, have an extensive and detailed knowledge of the temporal regularities of speech and actually apply this knowledge rapidly and unknowingly in optimizing their chance of correct decisions on perceived vowel length. This strategy for disambiguation of vowel length may seem an extremely complex and even cumbersome machinery for the communication of a very simple binary contrast. However, given the complexity of the temporal organization of speech, decision strategies in perception have to be complex in order to be efficient.

#### References

- Green, D.M. and J.A. Swets (1966): Signal detection theory and psychophysics, New York: Wiley.
- Lehiste, I. (1970): Suprasegmentals, Cambridge, Massachusetts, and London, England: M.I.T. Press.
- Nooteboom, S.G. (1972): "The interaction of some intra-syllable and extra-syllable factors on syllable nucleus durations", Institute for Perception Research Annual Progress Report 7, 30-39.
- Nooteboom, S.G. (1973): "The perceptual reality of some prosodic durations", JPh 1, 25-45.
- Nooteboom, S.G. and G.J.N. Doodeman (1978): "Perception of vowel length in spoken sentences", submitted for publication.

PREDICTING SEGMENT DURATIONS IN TERMS OF A GESTURE THEORY  
OF SPEECH PRODUCTION

S.E.G. Öhman, S. Zetterlund, L. Nordstrand and O. Engstrand,  
Dept. of Linguistics, Uppsala University, Sweden

Theory

We take the basic object of phonetic investigation to be the concrete sound gestalt produced by a speaker in a speech act, i.e. the spoken sentence. And we take the basic problem of our research to be that of explaining the physical structure of the sentence considered, not merely as a complex sound, but as a complex sound used as a vehicle for linguistic communication, or, briefly speaking, we take the problem to be that of explaining the phonetic structure of the sentence.

We explain the sentence phonetically by giving an explicit account of its linguistically functional, physical components, (atomic and compound), and of the methods by which these components are made to form a whole.

An oscillographic or spectrographic record of a spoken sentence does not by itself explain the phonetic structure of the sentence. In such a record, both the (primary) acoustic effects intended by the speaker to form the linguistically functional (atomic or compound) parts of the sentence, and the (secondary) acoustic traces of the speaker's efforts to bring the intended acoustic effects about, will be visible. As a first step in the phonetic explanation of a sentence we therefore try, on the basis of experiment, to distinguish the former acoustic effects from the latter, to define the intended (primary) effects in acoustic terms, and to explain with reference to the physiological properties of the organs of sound production and perception, why the speaker chooses to bring the intended acoustic effects about in just the way he does. In particular, we require detailed explanations of why the (secondary) acoustic traces of these efforts have the acoustic properties that they have.

As an example of this, consider a phonetic part of a sentence that has the form of a voiceless stop consonant such as [t]. It may be assumed that the intended acoustic effect in this case (the [t] per se) is the brief burst of friction noise. The formant transitions that follow this burst (into a following vowel), the

voiceless time interval that precedes it, and the formant transitions (from a preceding vowel) that precede this voiceless interval, are all to be regarded as secondary acoustic traces of the speaker's efforts to bring the burst about. These traces may be explained by showing that a burst of the type in question can only (or at least, most easily) be produced by building up air pressure behind an oral closure at a certain place and by then quickly releasing this pressure in the familiar manner. (A detailed analysis would of course have to make quantitative predictions on the basis of an explicit production and perception model.)

We evaluate an assumption about what does and what does not constitute the intended (primary) acoustic effects, in the total complex sound of a sentence, (1) on the basis of the possibility of explaining convincingly the detailed physical structure of the sentence (including secondary effects), given the physiological constraints on the production and perception mechanisms, and given that the speaker's goal is that of bringing about the assumed primary effects, and (2) on the basis of the depth of insight that the assumed phonetic structure gives us as regards the semantic function of the sentence in the speech act where it was used.

We do not assume that the phonetic structure of a sentence is segmental, nor that it is linear. Experience indicates, on the contrary, that the following picture is better justified:

In any language one operates with a finite inventory of types of atomic acoustic effects. We write  $E_1, \dots, E_n$  to denote these effect types for a given language (with  $n$  such types). Moreover, we use lower case letters (such as  $e$ ) to denote intended (atomic) acoustic effects, and we write

$$(1) \quad e \in E_1 \quad \text{or} \quad E_1(e)$$

to say that the atomic acoustic effect  $e$  is of type  $E_1$ .

Atomic acoustic effects may be combined to form larger acoustic units in either of two ways called coarticulation and sequencing. We write

$$(2) \quad e_1 + e_2$$

to indicate that the two atomic acoustic effects  $e_1$  and  $e_2$  are coarticulated, which means that they come about (are brought about) simultaneously, or, more accurately, that there is a point in time at which both these effects are heard. And we write

$$(3) \quad e_1 \bullet e_2$$

to indicate that the two atomic acoustic effects  $e_1$  and  $e_2$  are sequenced, which means that  $e_1$  comes about, whereupon  $e_2$  comes about as soon as can be done. It should be noted that, for all  $e_i, e_j$  and  $e_k$  the following equalities hold:

$$(4) \quad e_i + e_j = e_j + e_i$$

$$(5) \quad (e_i + e_j) + e_k = e_i + (e_j + e_k)$$

$$(6) \quad (e_i \bullet e_j) \bullet e_k = e_i \bullet (e_j \bullet e_k)$$

$$(7) \quad e_i + e_i = e_i \bullet e_i = e_i$$

Every language will have special rules according to which certain atomic acoustic effects (specific to that language) can be coarticulated and sequenced. These rules will also allow coarticulation and sequencing of nonatomic (compound) acoustic effects (differently in different languages). If  $e_1$  and/or  $e_2$  are compound effects, the expression  $e_1 \bullet e_2$  denotes the sequence in which  $e_2$  develops immediately after the last effect of  $e_1$  has emerged; and  $e_1 + e_2$  denotes an effect in which  $e_1$  and  $e_2$  develop simultaneously in such a way that the last effects of  $e_1$  and  $e_2$  coincide in time.

In most languages, we should expect to encounter compound acoustic effects of both of the following forms

$$(8) \quad (e_2 + e_5) \bullet e_3$$

$$(9) \quad e_2 + (e_5 \bullet e_3)$$

Here (8) is to be read: the compound effect in which  $e_2$  is coarticulated with  $e_5$ , immediately followed (as a whole) by  $e_3$ . And (9) is to be read: the compound effect in which  $e_2$  is coarticulated with a compound effect in which the effect  $e_5$  is immediately followed by  $e_3$ .

The linearity hypothesis, which we reject, excludes compound effects of the form (9) above.

The acoustic effects are related to articulation as follows. When the speaker says his sentence he knows what acoustic effects he intends to bring about and how they are to be arranged in terms of coarticulation and sequencing. In order to bring these effects about, he makes audible gestures with his organs of speech production, one gesture for each acoustic effect intended, whether atomic

or compound. I.e., the gestures can also be regarded as atomic or compound.

The gestures will be timed and executed in such a manner that (1) the intended acoustic effects come about and (2) no intended acoustic effects are destroyed by the bringing about of other effects.

We hypothesize that in a very considerable number of cases the segmental structure of sentences visible in oscillograms and sound spectrograms and, in particular, the temporal durations of these acoustic segments, can be explained as secondary effects due to the speaker's efforts to bring about the linguistically functional, primary acoustic effects.

The reasoning behind this hypothesis is, among others, this: The linguistically functional, intended acoustic effects are not, in general, required to have any particular duration. They are felt to be complete as soon as they are heard to emerge. A complex acoustic effect in which several atomic effects are coarticulated may, however, require for its execution a compound gesture one part of which is slower than all the others. If several of these gestures are started at about the same time, some of them may be completed earlier than the others in the sense that the effects that they aim at bringing about emerge before the others. To coarticulate all the effects, i.e. make them audible at the same time, the effects that emerge early will have to be maintained for some time while waiting for the remaining effects to materialize. Thus, acoustic segments with quasi-stationary qualities will arise not as a final end of the phonetic action but as a secondary consequence of the effort to reach a certain final end (the simultaneous sounding of the effects in question).

As an example of an alleged phonological contrast that seems eliminable on this paradigm we offer the Swedish contrast [vi:la] [vi:l:a] (rest, house) which we analyze as

v (stress + i) • l • a

v (stress + (i • l)) • l • a

where the stress effect which it takes a relatively long time to produce must be coarticulated with the vowel [i], (thus making the quickly producible [i] long) in the first case, whereas the stress effect is coarticulated with [i • l] in the second case (thus making the [i] long).

Among the acoustic effect types of most languages there will be certain (relative) pitch levels or compounds (sequences) of such levels. These pitch levels will in general be coarticulated with acoustic effects with the feature [+voice], especially vowels. We therefore expect that vowel duration will be strongly dependent on intonation in most languages.

In what follows some experimental data that have been collected to test this theory will be summarized.

#### Data

In two experiments (Zetterlund et al. 1978, Engstrand et al. 1978) we used a computer system (ILS) for manipulating prosodic parameters in natural speech to show that listeners consistently tend to overlook systematic durational variations in their identification of certain noun phrases such as compounds and lexicalized phrases. In an identification experiment we presented our informants with various synthesized versions of certain utterances (see Zetterlund et al. 1978, Engstrand et al. 1978) systematically changing fundamental frequency, vowel and consonant durations, and intensity. The responses were consistent to almost one hundred percent: The critical parameter for the listeners' identification of these utterances was F0. Although the acoustic analysis displays great variations in the duration and intensity parameters, our subjects apparently paid no attention to these potential cues in the presence of F0.

On the basis of the theory sketched earlier in this paper, we expect that a large F0 movement between two critical values would tend to space these points further apart in time than a small (or no) F0 change. To test this hypothesis we have in one experiment looked at words involving the Swedish word accent opposition. In bisyllabic accent 2 words in focus position F0 has to be low at the end of the first (stressed) vowel. The second vowel carries the sentence accent which is physically signaled as a high F0 at the beginning of the vowel. Consequently, most of the F0 rise has to take place during the intervening consonant occlusion. The corresponding accent 1 words do not display this upward shift but are characterized by a more or less level F0 contour during the consonant. The interesting thing about this is that the consonant in the accent 2 words seems to be significantly longer than the consonants in the corresponding accent 1 words.

It is known that F0 variations are accompanied by considerable vertical movements of the entire larynx box. Combined electro-myographic and larynx movement data that we have collected show that these vertical movements have definite muscular correlates, namely activity in the geniohyoid and sternohyoid muscles for upward and downward movement, respectively. Although the way the vertical movements mechanically affect the tension of the vocal folds is very much open to question, we can state that there is a very high positive correlation between larynx height and F0, and that F0 control involves a delicate coordination between small intrinsic musculature and larger supra- and infrahyoidal muscle masses. And, further, considering the comparatively large mass of the larynx box, it seems rather plausible to assume that its mechanical inertia in combination with the coordinative demands on the muscles should impose restrictions on the velocity with which it can conveniently and accurately be moved.

A further example is this: In an accent 1 word F0 is phonologically required to be low at the beginning of the stressed vowel. If the accent 1 word is preceded in the sentence by a relatively high F0, a downward movement of F0 is observed at the beginning of the accent 1 word sometimes extending into its stressed vowel. The duration of the consonant preceding the stressed vowel is found experimentally strongly to depend on the extent of the pitch drop through the beginning of the accent 1 word.

Finally, in order further to test the theory we have looked at the dispersion of the F0-values at various critical points and found that the standard deviations generally are extremely small, e.g. 2.6 Hz at the last peak of the hu segment in vita huset. We have several more examples of this kind.

Obviously, the most important task must be to establish the critical F0-values at different points in time with greater certainty. The perceptual tolerances that listeners have to deviations in the time-frequency domain should be investigated. We would also like to know what significance the exact shape of the F0 contour between the critical points have to a listener. As a matter of fact, the question whether the entire F0 contour or only some fixed values at certain line-up points relative to supra-glottal articulations are the intended and, therefore, phonologically crucial effects produced by the speaker is not yet completely



answered. Pilot experiments encourage us to believe that the latter hypothesis will prove to be true. This would mean, then, that a speaker is given a relatively large amount of freedom to choose ad hoc strategies for passing through the chain of successive critical F0 points. If this is true, a reasonable assumption is that the way transitions are brought about is adapted to fit some anatomical constraints on the larynx. Looking at our data we observe that the slopes of F0 rises and falls are characterized by constancy rather than variation.

#### References

- Engstrand, O., L. Nordstrand and S. Zetterlund: Experiments on the perceptual evaluation of prosodic parameters in compounds and lexicalized phrases. Paper given at The Phonetics Symposium held at the Department of Speech Communication, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, November 9-10, 1978.
- Zetterlund, S., L. Nordstrand and O. Engstrand: An experiment on the perceptual evaluation of prosodic parameters for word boundary decision in Swedish. Paper given at The Symposium on the Prosody of the Nordic Languages, Phonetics Laboratory, Department of General Linguistics, Lund University, June 14-16, 1978.
- Öhman, S.: Aktuell svensk forskning i fonetik. Tionde sammankomsten för svenskans beskrivning, Uppsala, april 1977. In: S. Eliasson, B. Loman, B. Sigurd, U. Telemann and S. Öhman: Svenskan i modern belysning. Fem översikter från Tionde sammankomsten för svenskans beskrivning (Ord och stil. Språkvårdssamfundets skrifter 9.) Lund: Studentlitteratur.

## MOTOR CONTROL OF SPEECH GESTURES

## Summary of Moderator's Introduction

James Lubker, University of Stockholm

Speech production theory is currently faced with several closely related and quite crucial issues which are well illustrated by the papers in this Symposium on Motor Control of Speech Gestures.

Perhaps central to these issues is the growing impatience among many phoneticians with what they see as a constraint to bend or adapt physiological/mechanical "fact" from motor control research to fit abstract linguistic constructs. This issue has been discussed in detail by a number of authors (e.g., Moll, et al., 1977; Fowler, et al., in press) and its general importance is reflected by the fact that it is taken up not only at this motor control symposium but in other papers (see MacNeilage's Status Report on Speech Production) and symposia at this IXth International Congress of Phonetic Sciences. Very briefly, the issue may be summarized as follows. Many investigators today contend that concepts which are relevant to the motor control of coordinated movements in general, whether from the walking movements of the hind leg of a cat or from arm movements about the elbow of a human being, are relevant also to the understanding of the motor control of the articulators for speech. It is argued that concepts related to the fine motor control of non-speech behaviors can and should be incorporated into speech production/motor control theory. In fact, I suspect that most investigators would accept such an argument, at least up to some specific point. That is, while many would agree that much fine motor control data from non-speech and from non-human research is of importance to speech production theory, they would also argue that in the end speech and language are distinctly human behaviors (although see MacNeilage's Status Report at this congress) and that the motor control of those behaviors is therefore unique, at least in some respects. For example, Bladon in his paper in this symposium, takes the view that "the physical facts of phonetics are at their most interesting when they serve to explain some aspect of phonology, to answer the question of why the sound systems of languages are the way they are." It is at this point that the impatience of many phoneticians becomes most evident, when they

note that in virtually all physiological/mechanical experiments on motor control mechanisms, correlates of abstract linguistic segmental units are conspicuous via their absence. Such units have proven extremely difficult to quantify. Thus, the question arises: should production theorists develop their own units and concepts which are based on actual experimental observations of motor control mechanisms in general and which are unbiased by notions and abstract concepts borrowed from linguistic theory? In the consideration of this question, either explicitly or implicitly, related questions and issues quickly arise. For example, Turvey and his associates (see, e.g. Fowler, et al., in press, for a review) describe much of modern phonetics research in production theory as consisting of "translation theories" designed to discover or elucidate the rules which could serve to translate from abstract linguistic units to the more concrete neurophysiological/mechanical data of speech motor control research. Turvey's use of Action Theory (Bernstein, 1967; Turvey, et al., 1978) and his development of the concept of "Coordinative Structures" represents an attempt to avoid such translation theories while at the same time not reject out of hand the use of all traditional linguistic concepts. The paper by Gay and Turvey in the present symposium provides some experimental consideration and discussion of the coordinative structures concept in speech motor control.

By its very nature, research in speech motor control, as exemplified by the reports in this symposium, is integral to issues such as these. Sussman, for example, discusses single motor unit behaviors and the insights they provide to temporal reorganization in coarticulation and to such prosodic events as stress, thus suggesting a means to provide "sensitive indicants of higher level linguistic conditions". Hirose provides data relevant to relationships between electromyographic activity and subsequent articulator movement. As MacNeilage points out in his status report paper at this congress, issues such as these cause questions concerning the role of feedback or closed loop control to become crucial. Indeed, the majority of the papers in this symposium at least refer to problems of feedback mechanisms while several specifically address themselves to such problems. Bladon proposes a "coarticulation resistance compiler" which is "linked

ambidirectionally" to satellite units in the motor control system. Abbs suggests a preliminary "multi-level control model" to account for observations of speech motor equivalence and compensatory articulation behaviors. Folkins also addresses the problem of motor equivalence, "functional interchangeability of activity level in different muscles", and compensations for mechanical modifications of articulator positioning. Perkell provides a discussion of recent compensatory articulation, or "bite-block", experimentation and thus the role of various sorts of feedback in speech motor control. He presents an example of the use of data from non-speech behaviors and in addition concludes that ideas such as those raised in motor control research are "closely related to questions about the nature of fundamental units which underlie the programming of speech production."

Thus, these papers on the Motor Control of Speech Gestures can be seen to confront some basic and crucial issues in phonetic theory. Further discussion of these and related issues is certain to bring us closer to an understanding of how it is that speech is generated and controlled.

#### References

- Bernstein, N. (1967): The coordination and regulation of movements, London: Pergamon Press.
- Fowler, C. A., P. Rubin, R. E. Remez, and M. T. Turvey (1978): "Implications for speech production of a general theory of action". In Language production, B. Butterworth (ed.), New York: Academic Press. (In press).
- Moll, K. L., G. N. Zimmerman and A. Smith (1977): "The study of speech production as a human neuromotor system" in Dynamic aspects of speech production, M. Sawashima and F. S. Cooper (eds.), 404-408, Tokyo: University of Tokyo Press.
- Turvey, M. T., R. E. Shaw, and W. Mace (1978): "Issues in the theory of action". In Attention and performance VII, J. Requin (ed.), Hillsdale, N. J.: Erlbaum (In press).

## SPEECH MOTOR EQUIVALENCE: THE NEED FOR A MULTI-LEVEL CONTROL MODEL

James H. Abbs, Speech Motor Control Labs., University of Wisconsin, 1500 Highland Ave., Madison, Wisconsin 53706, USA

In the last ten years it has become increasingly difficult to view the neuromotor execution of speech as a series of descending motor commands, reflecting, in some direct manner, an underlying matrix of phonetic features. Rather, it appears that patterns of speech muscle activity may depend upon moment-to-moment peripheral conditions and adaptive modification of descending commands at several nervous system levels. In the present paper I would like to outline some current thoughts with regard to these speech motor processes, including some data from our laboratory and a preliminary model to account for recent observations.

If one considers speech motor control teleologically, the adaptive modification and adjustment of descending speech motor commands, based upon peripheral feedback, is quite appealing. For example, the orofacial system obviously serves the multiple functions of chewing, swallowing, and breathing, in addition to speech. Other less natural intrusions include cigarettes between the lips, chewing tobacco, a pipe between the teeth, etc. Because many of these activities can be performed simultaneously, without major interference or conscious compensation, a nervous system capability for on-line adaptive adjustment appears almost necessary. Such semi-automatic adaptation also seems likely in laryngeal and respiratory control as well. Recent physiological investigations of the laryngeal and respiratory systems, as well as consideration of their anatomy, indicate the profound influence that torso, head, and arm movements have upon the specific muscle contractions required for speech. Trained singers are quite aware of these influences. However, for many speaking situations, we have little difficulty sustaining continuous and intelligible speech concurrently with vigorous body movements. Observing a physical fitness teacher perform calisthenics and at the same time continuously cohort his or her pupils is an obvious example of this phenomenon. A preferable observation might be a cheerleader at a U.S. football game. In these and other similar cases, e.g., a vigorous university lecturer (an example suggested by Peter MacNeilage), one is impressed with our ability to produce continuous speech without major interference. Possibly these multiple concurrent motor

programs could be generated and pre-adjusted in parallel, but such an organization seems contradictory to the obvious availability of multiple afferent monitoring channels, documented differences in their nervous system origins, the principle of economy, and current information on normal and abnormal speech motor control.

In part these observations can be explained by the provocative model offered by MacNeilage (1970). He suggested that speech motor commands are adjusted to assure that individual articulators reach semi-invariant target positions, despite a substantial degree of variability in their starting positions. This kind of compensatory capability was referred to by Hebb (1949) as motor equivalence, although Hebb's definition was not quite so restrictive. Since MacNeilage's original paper, experimental observations have extended our appreciation of the adjustment capabilities operating in the speech motor control system. These recent observations appear to require an expansion of MacNeilage's insightful model and support the operation of motor equivalence in its most encompassing terms.

#### Indirect Experimental Evidence

The hypothesized operation of motor equivalence adjustments to descending speech motor commands implies a repetition-to-repetition flexibility in the way that a particular speech utterance is generated. Recent investigations support the operation of such flexibility both with regard to trade-offs between individual articulators and between individual synergistic muscles acting to move the same articulator. For example, it has been shown that the upper lip, lower lip, and jaw trade off reciprocally in their cooperative contributions to oral opening, viz., when the jaw had relatively large displacements the upper and lower lips had relatively small displacements, and conversely (Abbs and Netsell, 1973; Hughes and Abbs, 1976; Watkin and Fromm, 1978). Other investigators (Hasegawa et al., 1976) have reported lip and jaw reciprocity not only in regard to displacement, but for lip and jaw velocities as well. The trade-offs reported in these studies were observed for multiple repetitions of the same utterance where the net contributions of the individual movements (i.e., total oral opening or net velocity of closing) was relatively consistent. Comparable analyses of speech lung volume control illustrate a similar pattern of reciprocal trade-off between movements of the abdomen and thorax in producing subglottal air pressures (Hixon et al., 1973). In our

laboratory we have found other patterns of articulatory trade-off, including reciprocal interactions between the tongue and jaw (Chuang and Abbs, In Progress).<sup>1</sup>

Not only do individual articulators appear to vary in their repetition-to-repetition contributions to a particular vocal tract objective, individual muscles appear to vary reciprocally in their combined contributions to an individual articulatory movement as well. In a recent experiment (Abbs and Kennedy, In Preparation), we found a reciprocal trade-off between the mentalis (MTL) and orbicularis oris inferior (OOI) muscles during repeated speech-related movements of the lower lip. This is in repetitions where the magnitude of MTL-EMG was relatively small, the magnitude of OOI-EMG was relatively large, and conversely. The flexibility of these adaptive speech motor command adjustments can be illustrated by considering this finding in relation to an earlier report by Sussman et al. (1973) of a parallel reciprocal trade-off between MTL-EMG magnitude and jaw lowering.

Overall these observations suggest that there may be several levels of programming and adjustment in the motor generation of speech. At some level, possibly corresponding to the phonetic feature input to the speech control system, overall vocal tract goals must be specified. However, due to the contrast between (1) the relative consistency with which these overall vocal tract goals are achieved, and (2) the variability of individual articulatory movements and muscle activity patterns, it would appear that these different output parameters are not programmed at the same levels of the nervous system.

#### Some Direct Evidence

The major issues with regard to this hypothesized motor control process concern the levels of the nervous system at which the adjustments might occur and the extent to which afferent feedback plays an important role. In attempts to more directly address these issues, several investigators have introduced unanticipated disturbances to the lips and jaw during ongoing speech (Bauer, 1974;

-----  
 (1) These patterns are most apparent in phonetically naive speakers. "Trained phoneticians" appear to produce speech, especially with regard to these reciprocal articulatory movements, quite different from that of normal speakers (cf. Gay, 1976).

Folkens and Abbs, 1975; 1976; Kennedy, 1977; Murphy and Abbs, In Progress). In these studies it was reasoned that if the nervous system sites where adaptive adjustments occurred were at "lower levels", semi-automatic, short-latency compensations would prevent unanticipated disturbances from interfering with ongoing speech. In the 15 subjects run with this particular paradigm, there have been no cases of disruption to ongoing articulation. In those studies where the latency of the compensations was discernible, it ranged from 25-50 msec. Compensatory responses have been observed in the muscles of the articulator to which the load was applied as well as in other articulators contributing to the same vocal tract goal, i.e., loads applied to the jaw yielded compensations in both the upper lip and lower lip musculature. The diffuse yet functional nature of these multiple compensatory responses corroborates the earlier suggestion that individual articulators and individual muscles can be adjusted flexibly to achieve desired overall vocal tract objectives. Based upon these findings, it appears that lower levels of the nervous system may be plausible sites for the adaptive modification of descending motor commands. Lower level sites for these adjustments are supported also by the observation that while the subjects in these studies perceived the articulator loading, they were unaware of generating the compensatory adjustments.<sup>2</sup>

A recent finding that might point to the possible origin of these compensatory adjustments is the observation that individuals with cerebellar disease and ataxic dysarthria are unable, without practice and conscious intervention, to adjust their lip movements to overcome experimental stabilization of the jaw (Netsell et al., In Preparation). Indeed, these patients report that many of their speech movements must be "consciously controlled". Certainly, if one accepts Eccles' (1973) suggestion, the cerebellum, with its multiple afferent and efferent connections, would be a primary candidate for yielding the semi-automatic, unconscious adjustments apparently required for normal speech. Other yet lower level sites

-----  
 (2) In our experience, the major difficulty in these unanticipated disturbance studies is discerning the peripheral manifestations of the disturbance. That is, while ongoing speech is seldom disrupted, the compensatory degrees of freedom are so great that one cannot always ascertain which muscles or movements were involved. This problem has apparently impeded some investigations using aerodynamic disturbances (Perkell, 1976).

might include areas in the brain stem where single point electrical stimulation yields very complex and semi-coordinated gestures of the laryngeal, masticatory, lingual, and facial musculature (Luschei, Personal Communication).

#### A Preliminary Model

Figure 1 is a schematic attempt to represent the motor control processes warranted from the data cited above.

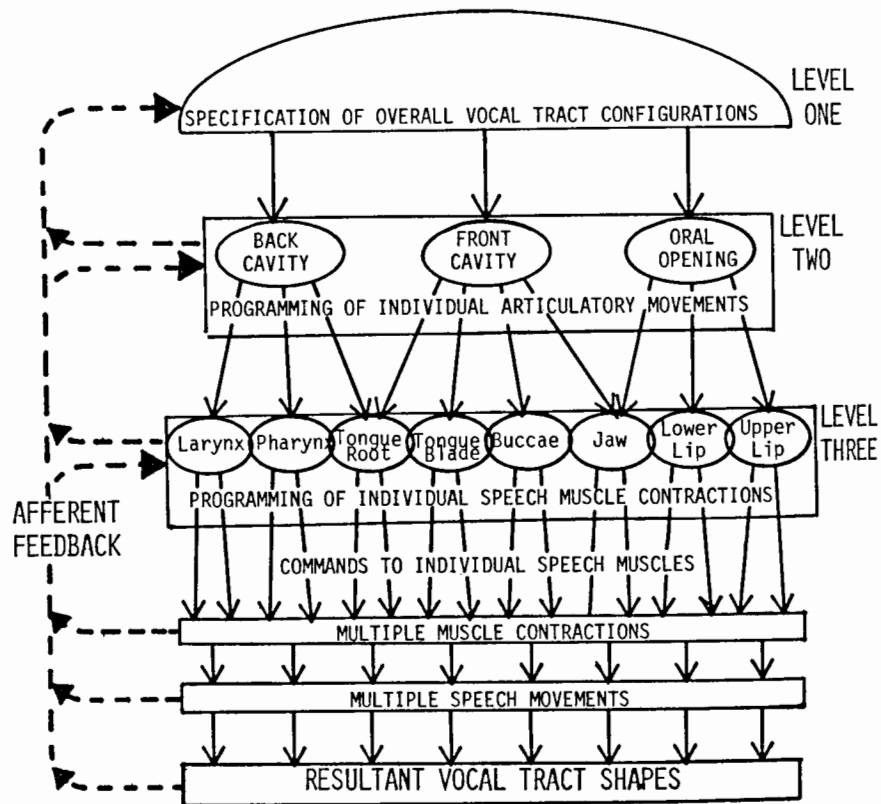


Figure 1

A multi-level model of the speech motor programming process. Solid lines represent descending control signals and dashed lines afferent feedback information. Dashed lines between levels of programming represent the ascending components of internal feedback pathways.

This model posits three levels of speech motor programming. At the highest level, overall vocal tract goals are specified, perhaps corresponding to some matrix of phonetic features. At the least, these goals represent the temporal-spatial configurations necessary for appropriate modulation of aerodynamic and acoustic signals. The second level of programming is involved in determining the particular set of individual movements that are to be employed in achieving the desired vocal tract goals. The third and final level of programming is concerned with specifying the individual muscle contraction patterns necessary to the generation of individual articulatory movements. These two lower levels of programming are based upon the observations (cited earlier) that (1) individual articulatory movements are not invariant with regard to particular vocal tract goals, i.e., repetitions of the same speech element, even if acoustically and perceptually similar, are produced often by different combinations of articulatory movements, and (2) individual muscle contractions are not invariant with regard to particular articulatory movements, i.e., repetitions of an articulatory movement are produced by different combinations of individual muscle contractions. As shown in Figure 1, it is posited also that the programming/adjustment of descending motor commands is accomplished with the aid of afferent feedback. This feature of the model is based upon the observations of compensatory responses to unanticipated articulator loading. That is, while it is plausible to consider parallel pre-adjustment of multiple motor commands (through some sort of efferent copy), in response to steady-state, anticipated disturbances (Lindblom et al., In Press), rapid adjustments to dynamic, unanticipated loads appear to require an afferent feedback control capability.

It is apparent from this representation that the model previously offered by MacNeilage does not account for all the motor command adjustments that apparently are accomplished by the speech motor execution system. That is, the adjustments to descending motor commands, at least as evidenced by the data cited above, obviously involve more than compensations for variations in individual articulator starting positions. Indeed, it appears that the primary controlled output parameters of the speech production system are not individual articulatory movements, but a series of overall vocal tract configurations. This model has other implications as

well. For example, analyses of individual articulatory movements or muscle contractions in relation to underlying phonetic features appear to be based upon the assumption that there is but a single level of speech motor programming. However, with multiple levels of adjustment, there is some question as to whether individual muscle contractions or articulatory movements are related, except in a probabilistic manner, to overall vocal tract phonetic features. If such a direct relationship exists, it may be necessary to hypothesize different features or to reallocate the current features, at least in part, to lower levels of the nervous system.

#### References

- Abbs, J. and R. Netsell (1973): "Coordination of the jaw and lower lip during speech production", ASHA Convention, Detroit.
- Bauer, L. (1974): "Peripheral control and mechanical properties of the lips during speech", M.S. Thesis, Univ. of Wisc., Madison.
- Eccles, J. (1973): The Understanding of the Brain, New York: McGraw.
- Folkins, J. and J. Abbs (1975): "Lip and jaw motor control during speech", JSHR 19, 207-220.
- Folkins, J. and J. Abbs (1976): "Additional observations on responses to resistive loading of the jaw", JSHR 19, 820-821.
- Gay, T. (1977): "Cine and EMG studies of articulatory organization", in Dynamic Aspects of Speech Production, M. Sawashima and F. Cooper (eds.), 85-102, Tokyo: Univ. of Tokyo Press.
- Hasegawa, A., M. McCutcheon, M. Wolf and S. Fletcher (1976): "Lip and jaw coordination during the production of /f,v/ in English", JASA, S84, 59.
- Hebb, D. (1949): The Organization of Behavior, New York: Wiley.
- Hixon, T., M. Goldman and J. Mead (1973): "Kinematics of the chest wall during speech production", JSHR 16, 78-115.
- Hughes, O. and J. Abbs (1976): "Labial-mandibular coordination in the production of speech", Phonetica 33, 199-221.
- Kennedy, J. (1977): "Compensatory responses of the labial musculature to unanticipated disruption of articulation", Ph.D. Thesis, Univ. of Washington, Seattle.
- Lindblom, B., J. Lubker and T. Gay (In Press): "Formant frequencies of some fixed mandible vowels and a model of speech motor programming by predictive simulation", JPh.
- MacNeilage, P. (1970): "The motor control of serial ordering in speech", Psych.Rev. 77, 182-196.
- Perkell, J. (1976): "Response to an unexpected suddenly induced change in the state of the vocal tract", MIT Res.Lab.Elect. 117, 273-281.
- Sussman, H., P. MacNeilage and R. Hanson (1973): "Labial and mandibular dynamics during the production of bilabial consonants", JSHR 16.
- Watkin, K. and D. Fromm (1978): "The control of labial movements by children", ASHA Convention, San Francisco.

## MOTOR CONTROL OF COARTICULATION: LINGUISTIC CONSIDERATIONS

R.A.W. Bladon, Department of Linguistics, University College of North Wales, Bangor, U.K.

Various orientations to the motor control of speech

An orientation advocated recently by Moll, Zimmermann and Smith (1977) calls for priority to be given, in research on speech motor control, to exclusively neurophysiologically-based studies. The need is, they argue, to determine the properties of the human neuromotor system based on investigations of movement, muscle contraction and motor unit activity, freed from any constraints or a priori constructs imposed by linguistic considerations. Any processes or units of neuromotor coding which such enquiry were to establish might or might not subsequently turn out to correlate with linguistic units such as the phone, the feature or the syllable.

Without wishing to deny that the approach of Moll et al. has value, we propose to offer to this symposium the opposite orientation, wherein aspects of the descriptive linguistic apparatus are of prime importance. This decision reflects partly the conviction of a linguistic phonetician that the physical facts of phonetics are at their most interesting when they serve to explain some aspects of phonology, to answer the question why the sound systems of human languages are the way they are. The decision is also derived from the evidence that a wide range of phenomena of coarticulation are not obviously explainable (as yet, at least) in terms of the neuromotor system such as motor unit activity or articulatory velocity and inertia, but are referable to linguistically-defined entities which they thereby can validate.

In addition, many models of the speech production processes seem to occupy the middle ground between these two extreme positions. Among the proposals which might be grouped together here are those of Kozhevnikov and Chistovich (1965), Henke (1966), MacNeilage (1970), Gay (1977a) and Perkell (1977). In very general terms, these models are of a basic "wedding-cake" form such as Figure 1: that is, they are arranged sequentially as tiered boxes, with a distinct top and bottom corresponding to mechanisms associated respectively with more central cortical functions and with more peripheral ones. The number of tiers, and the content of each one, is stylised and is not meant to be attributed specifically



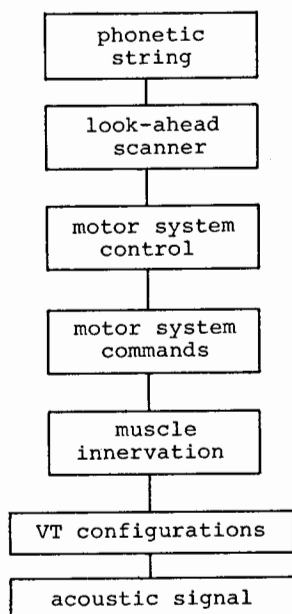


Figure 1

Typical "wedding-cake" structure of some speech production models

control system for coarticulation is organised not in a unidirectional, tiered fashion, but as a set of satellites (many representing linguistic factors) linked ambidirectionally to a nuclear body, the CR (Coarticulation Resistance) compiler. The following formulation is a brief summary of this theoretical position. Justification of it, of an organisation which reduces the emphasis on the supposed sequential nature of the speech processors and which minimises the "more central/less central" distinction, and of the particular components recognised in the model, has been given elsewhere (Bladon and Al-Bamerni, 1976; Bladon, 1978).

Information relevant to the direction and domain of coarticulatory effects seems to derive from a wide variety of satellite sources, some quasi-universal (such as the boundary of an intonation-group, which is widely observed to impede the temporal spread of coarticulated features), some language-specific (such as the

to any author. Varying types of feedback between tiers are postulated, but not shown in the figure. A possible objection to the sequential arrangement is that, while no doubt well motivated for the lowest tiers representing the transduction of speech between various fairly accessible transmission channels, such an arrangement is more speculative as a claim about the higher levels of the central nervous system.

The "wedding-cake" models share with our own orientation (and in contrast to Moll et al.) an interest in the linguistic nature of the input, which they normally state to be a string of discrete phonological units.

By contrast with the claims implied by Figure 1, however, our conception of the "upstream" processes feeding data to the motor

report by Ladefoged (1967) that while French and English both show a /k/ coarticulatorily advanced before an /i/ vowel, only French shows the similar effect after /i/), and some speaker-sensitive. Our discussions of these factors have led Kent and Minifie (1977, 120) to write: "Perhaps the solution to coarticulation is as complex as this multiplicity of factors suggests, but ... (they) represent the contributions of many unknown, or poorly known effects". This comment is valid; but it is not a criticism of our position. It seems to be inescapable that the control of coarticulation in speech is indeed governed by a multiplicity of factors.

With a view to integrating these disparate factors in a theory, let us initially postulate the notion of coarticulation resistance (CR) as the central principle of articulatory control. The speech production mechanism is hypothesised to have continuous access to CR information, which can be considered to attach to each allophone and phonetic boundary. It is important to realise why an initial CR specification is tabulated for each allophone. A classic demonstration of this is afforded by the RP English /l/ allophones, of which dark syllabic [ɫ] ('fiddle') is highly resistant to coarticulation, dark nonsyllabic [ɫ] ('feel') is somewhat less so, and clear nonsyllabic [l] ('leaf') is very much less resistant. We are aware of no explanation of this behaviour in any terms, linguistic or neurophysiological: the idiosyncrasies can only be handled by assuming something like an allophone-specific assignment of a CR value. This numerical value is re-computed at a level of articulatory planning by the CR compiler to take account of the wide range of relevant coarticulatory constraints.

In what follows we presuppose a phonological apparatus broadly of the kind of Chomsky and Halle (1968). Within their phonological component, various kinds of linguistic construct will be examined, in conjunction with the control mechanisms related to them.

#### Phonological constructs related to units of articulatory planning

Many early coarticulation studies hoped to identify a single determinant regulating the control of the domain of coarticulatory effects: an invariant unit of production which would have a linguistic counterpart, such as the phone, syllable or phonetic feature. The hope was a vain one. The weight of evidence now available suggests that the coarticulatory control mechanism is sensitive not to any one invariant unit alone but, at different times, to (at least) all those three.

The phonetic feature is at the basis of coarticulation theory in that typical cases of coarticulation arise by definition from the asynchrony of events associated with different articulators. This is reflected phonetically in the temporal spreading of a feature. It is true that the speech control mechanism can be highly sensitive to the feature being coarticulated. Thus, for example, it has been shown that English /s/ occurring in CCC clusters blocks the spread of anticipatory jaw-opening before /æ/ (Amerman, Daniloﬀ and Moll, 1970); and that /s/ resists any shift in its tongue-bladeness (towards a tip articulation) adjacent to /t d n l/ (Bladon and Nolan, 1977); but that this resistance to coarticulate is specific to the coarticulated feature in question, because /s/ freely allows coarticulated labialisation anticipating an /u/ vowel (Daniloﬀ and Moll, 1968). It is equally true that to propose the feature as the sole unit of coarticulatory control would be unattractive, as it would not account for example for British English clear [ɪ], which is quite free in its coarticulation in respect of any of the features vowel-quality, lateral-quality and voicelessness indiscriminately (Bladon and Al-Bamerni, 1976).

Numerous cases, such as the last-mentioned, argue for the phoneme (or perhaps better, the extrinsic allophone) as the unit of articulatory planning. Two further examples may be mentioned. In Italian, intervocalic consonants demonstrate an equal degree of coarticulated tongue-body movement with both a preceding and a following vowel, thus irrespective of syllable boundaries. In French, the anticipatory spread of velum lowering before a nasal, as revealed by EMG, is over a limited domain within a string of preceding oral vowels (Bladon and Carbonaro, 1978; Benguerel et al., 1977). Such arguments for the allophone-sized unit tend, however, to be of a "default" kind, postulated whenever coarticulation fails to coincide with syllable boundaries in some sense. Generalising the allophone, in the interests of proposing an invariant unit, to cases which the syllable could have successfully delimited, has led to an overall too weak hypothesis concerning coarticulatory domain, such as that of Henke's model (1966), which predicted coarticulatory activity whenever a segment showed no antagonistic specification. Our model avoids this problem by two expedients: first, by a segment-specific index of CR (referred to earlier) which inhibits coarticulatory spread in appropriate circumstances, and second, by recognising a plurality of articulatorily-relevant

units which will include the syllable as required.

The phonological syllable, neglected by Chomsky and Halle, has since 1968 enjoyed a revival. Syllable-structure rules in phonology would define the syllable differently for different languages; nevertheless, the structure CV has a claim to universal preference in that, first, there appear to be no languages without CV syllables, second, several languages have syllables of only the CV type, and third, CV is the attested structure in early language acquisition. We profoundly disagree, therefore, with Gay's opinion (1977a) that Kozhevnikov and Chistovich's notion of an articulatory syllable of the form  $C_0V$  (where  $C_0$  stands for any number of consonants) is "an unnatural and counterintuitive syllable that bears no simple correspondence to common linguistic or phonetic units." Within their articulatory syllable, it will be recalled, coarticulation was hypothesised to be maximal. A great deal of evidence supports this hypothesis, notably the labialization of a string of C before /u/ in Russian (Kozhevnikov and Chistovich, 1965), in English (Daniloﬀ and Moll, 1968) and in French (Benguerel and Cowan, 1974); and also the finding (Bladon, 1977) that even the relatively weaker lip-rounding accompanying English /r/ extended leftwards to the same  $C_0V$  boundary.

#### Other substantive constructs in phonology

Explanation of the control of coarticulatory behaviour in VC positions has remained elusive. Relevant data here include the anticipatory nasalisation of English vowels before nasals (Moll and Daniloﬀ, 1971); American English /r/ which coarticulates with adjacent vowel quality more readily in the final position than in the initial CV position (Lehiste, 1964); or, in VCC sequences, the consonantal influence upon tongue apex position in V (Amerman and Daniloﬀ, 1977). Current phonological theory suggests an explanation in terms of the phonological strength hierarchy. Based on a variety of evidence including sound-change, phonological segments, sequences and positions in the word are assigned a degree of phonological strength. VC positions are weak, since they show more phonological assimilations and elisions. It is reasonable to suggest that the coarticulatory control mechanism is sensitive to this, as to other linguistic properties.

A second such property is the lexical representation of the inventory of phonological items in a language. The degree to which a lateral, for instance, undergoes vowel-quality coarticulation

varies according to the number of laterals in a language's phonological system: in our data, Irish, with three laterals to be kept distinct, shows very little quality coarticulation in comparison with American English, with only one (but highly coarticulated) lateral; Swedish or Italian, with two laterals each, fall in between with respect to coarticulation. The need in such cases to maintain phonemic distinctions (short of the point of incipient sound change and phonemic restructuring) has widely been held to place an upper bound on the extent of coarticulatory behaviour. The principle appeared to make the wrong predictions in the data of Benguerel and Cowan (1974), however, who found that lip protrusion anticipating French /u/ could sometimes extend transconsonantly into the preceding vowel, despite the apparent threat to the lexical contrast /i - y/ in French.

It seems certain that rapid-speech variations are subject to a degree of coarticulatory control. Gay (1977b) showed that at a fast speaking rate a vowel F2 transition effectively begins at a point of greater overlap with the preceding consonant than at a normal speaking rate. Rapid or casual speech variants are coming under scrutiny by phonologists in order to validate their substantive hypotheses of rule ordering. In the derivation of the (ultimate?) rapid-speech form [də.viː] 'divinity', Stampe (1972) demonstrates fairly convincingly that phonological processes do not apply in a linear order, but whenever the configurations they would eliminate arise. Among the processes concerned is the coarticulatory one of intra-syllable vowel nasalization, which re-applies three times, as successively more rapid forms are derived. This cyclic manner of application has important implications for the operational design of the motor control component of the speech production processes, and strongly supports the notion of ambidirectional, on-line exchange of information between the CR compiler (or its equivalent) and the linguistic rule system.

The testing of these various elements of the coarticulatory model by predicting from them onto new data, turns out to be partly successful, but, as has been demonstrated for several cases, partly unsuccessful. Apparently, no one linguistically-related mechanism will explain all or even a majority of observed coarticulatory behaviour. How to assign a weighting to the separate contribution of each mechanism, and indeed how many such mechanisms there are, remain research questions for the future.

### References

- Amerman, J.D. and R.G. Daniloff (1977): "Aspects of lingual coarticulation", *JPh* 5, 107-114.
- Amerman, J.D., R.G. Daniloff and K. Moll (1970): "Lip and jaw coarticulation for the phoneme /æ/", *JSHR* 13, 147-161.
- Benguerel, A.-P. and H.A. Cowan (1974): "Coarticulation of upper lip protrusion in French", *Phonetica* 30, 41-55.
- Benguerel, A.-P., H. Hirose, M. Sawashima and T. Ushijima (1977): "Velar coarticulation in French", *JPh* 5, 159-168.
- Bladon, R.A.W. (1978): "Some control components of a speech production model", in *Current Issues in the Phonetic Sciences*.
- Bladon, R.A.W. and A. Al-Bamerni (1976): "Coarticulation resistance in English /l/", *JPh* 4, 137-150.
- Bladon, R.A.W. and E. Carbonaro (1978): "Lateral consonants in Italian", *Italian Linguistics*, forthcoming.
- Bladon, R.A.W. and F.J. Nolan (1977): "A videofluorographic investigation of tip and blade alveolars in English", *JPh* 5, 185-193.
- Chomsky, N. and M. Halle (1968): *The Sound Pattern of English*, New York: Harper & Row.
- Daniloff, R.G. and K. Moll (1968): "Coarticulation of lip-rounding", *JSHR* 11, 707-721.
- Gay, T. (1977a): "Articulatory units: segments or syllables?", *Paper read at Symposium*, Boulder, Colorado.
- Gay, T. (1977b): "Articulatory movements in VCV sequences", *JASA* 62, 183-193.
- Henke, W. (1966): *Dynamic articulatory model of speech production using computer simulation*, Ph.D. thesis, M.I.T.
- Kent, R.D. and F.D. Minifie (1977): "Coarticulation in recent speech production models", *JPh* 5, 115-133.
- Kozhevnikov, V.A. and L.A. Chistovich (1965): *Speech: Articulation and Perception*, *JPRS* 30, 543, US Department of Commerce.
- Ladefoged, P. (1967): *Linguistic Phonetics*, UCLA Working Papers in Phonetics 6, Los Angeles: UCLA.
- Lehiste, I. (1964): *Acoustic characteristics of selected English consonants*, Bloomington: Indiana UP.
- MacNeilage, P.F. (1970): "Motor control of serial ordering of speech", *Psych.Rev.* 77, 182-196.
- Moll, K.L. and R.G. Daniloff (1971): "Investigation of the timing of velar movements in speech", *JASA* 50, 678-684.
- Moll, K.L., G.N. Zimmermann and A. Smith (1977): "The study of speech production as a human neuromotor system", in *Dynamic Aspects of Speech Production*, M. Sawashima and F.S. Cooper (eds.), Tokyo.
- Perkell, J. (1977): "Articulatory modeling...", in Carre R. et al. *Modèles articulatoires et phonétiques*, G.A.L.F., 197-192.
- Stampe, D. (1972): *A dissertation on natural phonology*, Ph.D. thesis, U. Chicago.

## SEGMENTAL INVARIANCE RECONSIDERED

R.G. Daniloff, Purdue University, W. Lafayette, Indiana, USA  
 M.A.A. Tatham, Languages and Linguistics, University of Essex,  
 Colchester, UK

Stop consonant articulation is context sensitive, Fromkin (1966). Such context sensitivity, often related to coarticulation, is taken as support for complex, high level "encoding", Kozhevnikov and Chistovich (1965). The duration and force of labial closure have been variously shown to be context sensitive to (1) syllable position, (2) voicing, (3) stress and (4) vocalic context. We concluded that previous cinefluorographic, electromyographic, and palatographic studies may have overestimated the extent of context sensitivity by failure to control for tempo, speech effort, task learning, and sentential context. The purpose of the study reported herein was to reassess the context sensitivity of the EMG impulse associated with labial-stop closure.

Procedures

4 adults served as subjects. Each was required to repeat previously tape recorded sentences as he heard them over earphones. The tape recorded sentences were carefully controlled for constant tempo, stress levels, and good phonetic quality. Speech tokens consisted of all combinations of  $C_xVC$ ,  $CVC_x$ ,  $C_xVC_x$ , 'VCV, and  $V'CV$  tokens,  $C_x = [p,b]$ ,  $C = [d]$ ,  $V = [i,u,\lambda]$ . CVC items were spoken in the sentence, "He'll spoof the [CVC] again", and VCV items in, "Smell this poof of [VCV] again", such that each token received primary, sentence level stress. Subjects visually monitored their vocal output as they spoke, keeping it at 60-65 dB, SPL. Each subject practiced the task of repeating the tape recorded sentences with as controlled a tempo, vocal output, and desired stress level as possible.

Bipolar silver disc surface electrodes on the upper lip served to detect the EMG pulse on the orbicularis oris for the lip closing gesture for [p,b]. The EMG and Raw Voice signals were recorded, rectified, and integrated, and on a single, 5 channel mingograph output, raw/integrated EMG, voice, and timing signals were displayed. The upper lip-surface electrode array was chosen because: (1) the upper lip is accessible, (2) labial surface EMG signals are quite interpretable-reproducible, (3) myo- and biomechanically

the lip in its closing gesture is a simple system, (4) the labial closure gesture has been extensively studied, (5) surface electrodes potentially yield a better estimate of whole-muscle activity than do needle/hooked wire electrodes, (6) labial closure is reputedly context sensitive.

Subjects received extensive practice beforehand; subjects who could not relax the lips to nearly zero-baseline EMG activity between utterances were rejected. Each of 40 tokens was repeated 26 times for a total of 1040 repeated, randomized sentences spoken at one sitting, with pauses every 10 minutes for relaxation.

Results:

The results of the study are based upon 4 criterion measures: the peak amplitude of the EMG pulse for lip closure, the duration of the EMG pulse, the delay between EMG onset and acoustic burst release, and the delay between peak EMG level and burst release. Data were analyzed using 3 way ANOVAs, with a conservative  $\alpha \leq 0.1$ . The concise results are shown in the table below. Interpretation of the results is based upon the following assumptions: lip closure for stops is mediated primarily by orbicularis oris contraction (00). M.O.O. force of contraction and the height of the integrated EMG signal are linearly, if not monotonically related. If one of the particular aspects of context-sensitivity being investigated were a part of speakers' linguistic competence, it was our expectation that all 4 speakers should show a statistically significant and similar shift in the labial closure criterion measure associated with that aspect of context since by our estimate, 25 repetitions of each token offered a firm basis for statistical inference.

The peak EMG amplitude measure presumably relates directly to maximum force of M.O.O. contraction. As shown, in every case, one or more subjects for one or more tokens showed a non-significant change in EMG peak amplitude; and in fact, for all 4 contextual effects: voicing, syllable position, vowel, stress - at least one subject showed a reversal of trend, with differences in EMG peak being just opposite those shown for 2 or 3 of the subjects. Thus, there is only a modest trend for voiceless stops in /i/ context to be modestly more effortful, muscularly. Stress and syllable position had no consistent effect upon peak EMG amplitude. Duration of the EMG pulse revealed a strong dependence upon context such

	VOICE	POSITION	VOWEL	STRESS
Peak EMG to Burst Onset	voiceless>voiced, moderate effect; subject, token dependent	voiced initial>voiced final; moderately strong effect; strong voicing, moderate subject dependence	not consistent, weak effect; strong token, subject dependence	V2 stress>V1 stress; moderate effect; small subject dependence
Delay EMG Onset to Burst	voiceless final>voiced final; moderate effect; token, position dependent	initial>final; strong effect; small token, subject dependence	/u/>/i/ weak effect; strongly token, subject dependent	V2 stress>V1 stress; strong effect; small subject dependence
Duration of EMG Pulse	-voice>voice; strong effect; little subject or token dependence	initial>final; strong effect; small subject, token dependence	not consistent; moderately weak effect; very token, subject dependent	V2 stress>V1 stress; strong effect; small subject dependence
Peak Amplitude EMG	-voice>voice; moderate effect; subject, token dependent	initial>final; weak effect; very subject, token dependent	/i/>/u/>/u/, moderate effect; subject, token dependent	not consistent; weak effect; strongly subject dependent

that voiceless stops were longer than voiced, initial stops were longer than syllable final stops, and pre-stress-position stops were longer than post-stress stops. In all three cases, tokens for one subject failed to achieve significance. In addition, for the voicing effect, one subject's voiced stops were significantly longer than his voiceless tokens. For the time delay between EMG onset and burst release, the effect of syllable position was fairly strong in that all initial stops began earlier and, in 6 of 8 cases, the difference was significant. For voicing, there was a modestly strong trend for voiceless stops to begin earlier, in 11 of 12 cases, with 10 of the 12 cases being significantly greater. Stress had a strong effect in that the pre-stress stop began earlier for all 4 subjects, significantly so in 3 of 4 cases. Vowels had no consistent effect upon delay. The peak EMG to acoustic release temporal delay measure shows a weak dependence upon voicing; the effect of the vowel upon delay was weak and inconsistent. The effect of stress upon this measure was only moderate in that for all subjects, pre-stress stops had earlier occurring peaks, but these differences were significant in only two of four subjects. The effect of position upon this delay measure was moderately strong in that in all 8 cases, syllable initial vowels had earlier occurring EMG peaks, and in 6 of 8 cases, the differences were significant.

#### Conclusions

Contrary to the work of Fromkin, syllable final stops were shorter in all cases, significantly so in 6/8 cases, than initial stops. Syllable position had no consistent effect upon the amount of muscle activity for closure, but initial stops began earlier, vis-à-vis onset or peak of EMG and burst release, in all cases, and significantly earlier in 12 of 16 cases. Thus, syllable initial stops are generally longer and earlier in onset, but not muscularly more effortful. Vowel context effects were enigmatic. It was expected that one would have earlier and stronger EMG pulses for [i] than for [ʌ] than for [u]. This was not the case; in the majority of cases, vowel context had non-significant effects on most EMG measures, and even when significant, the direction of the trend varied from subject to subject. Stress was potent as a factor such that pre-stressed stops began earlier in all cases, significantly so in 5 of 8 cases, and in 3 of 4 cases, the EMG



pulse was significantly longer. However, stress had no systematic effect upon the amount of muscle activity needed for closure. Finally, voice as a factor had no systematic, cross subject effect upon amount of muscle activity. With only one reversal (significant), in 9 of 12 cases, voiceless stops were longer than voiced stops, and EMG onset began earlier, vis-à-vis release, in 10 of 12 cases, one exception being a significant reversal. The effect of voice upon the EMG peak to burst release measure was highly variable.

The most startling result was that without exception, at least one subject, for at least one token showed either a non-significant context effect, or a reversal of trend for a given context effect. According to a strict criterion of all subjects and all tokens revealing a significant change in criterion measure, then, not one of the contextual effects investigated is less than idiosyncratic, i.e.: the contextual effects are a trend, but not absolutely a component of linguistic performance. Further analysis showed that subject sex and naïveté had no effect upon the results. Surprisingly, the phonetic shape of the syllable, e.g.: C<sub>x</sub>VC<sub>x</sub> vs. CVC<sub>x</sub> vs. VCV had a profound effect upon all criterion measures, and was probably the single most potent effect found in this study. It is difficult to explain why vowel context did not produce more, and more systematic changes in the size and timing of the EMG patterns for labial closure. It may be the case that coarticulation of stops and vowels, known to be quite a strong effect, is highly idiosyncratic. Or, it may be the case that the transformation of muscle activity into final vocal tract shape is complex, and non-linear, within and across subjects, so that interpretation and comparison of EMG data are more complex than is suspected. It is our conclusion that labial closure as an articulatory gesture is relatively context insensitive as far as amount of muscle activity is concerned. It is context sensitive as far as syllable position, voicing and stress are concerned in that voiceless, initial, pre-stress stops are generally longer and begin earlier: however, certain subjects and tokens violate this trend. We conclude that electromyographic signals, especially vis-à-vis coarticulation, may be more complex to interpret than is presently suspected.

#### References

- Fromkin, V.A. (1966): "Neuromuscular specification of linguistic units", L&S 9, 170-199.
- Kozhevnikov, V. and L. Chistovich (1965): Speech: articulation and perception, JPRS 30, 543, U.S. Bureau Commerce, Washington, D.C.

MASSETER, TEMPORALIS, AND MEDIAL PTERYGOID ACTIVITY WITH THE  
MANDIBLE FREE AND FIXED

John W. Folkins, The University of Iowa, Iowa City, Iowa, U.S.A.

Experiment One

Introduction

In general each of the speech articulators is acted on by a number of anatomically different muscles. The muscles which act on a given articulator may either 1) have activity levels which are functionally interchangeable or 2) each muscle may have a separate function which is seldom (if ever) accomplished by substitution of activity in other muscles. The Russian physiologist, Nicholas Bernstein (1967), has stressed the extent to which the first possibility; i.e., the functional interchangeability of activity level in different muscles, operates in normal human movements. In relation to speech, MacNeilage (1970) presents a perspective which (in this one respect) is related to Bernstein's ideas. MacNeilage believes there is a ubiquity of variability in muscle activity for attainment of vocal tract targets.

Textbooks of speech anatomy (Palmer, 1973; Zemlin, 1968) imply that masseter, temporalis, and medial pterygoid act in a similar manner to raise the mandible during speech. If this is the case, these muscles might typically operate interchangeably in many combinations for similar speech movements. However, the jaw closing muscles have been studied extensively in the dental literature (e.g., Kawamura, 1974) and the anthropological literature (e.g., Hylander, 1975). These studies have illustrated important differences in function and activity patterns between jaw-closing muscles. On the basis of these studies one might expect that each jaw-closing muscle has a specific functional role during speech and its activity is not typically interchanged with activity in other muscles.

There is not room to discuss the electromyographic (EMG) studies of the jaw-closing muscles during speech. However, the research to date does not provide much data concerning the above issue. Therefore, the purpose of the present study is to examine EMG activity and make comparisons between and within jaw-closing muscles during speech.

### Method

Hooked-wire electrodes were used to record EMG from masseter, temporalis, and medial pterygoid in four normal adults. Jaw movement was transduced with a strain gauge technique. Each subject produced three to six repetitions of 11 isolated syllables, seven syllables in a carrier phrase, trains of syllables at various rates, and the rainbow passage. The limitations on the length of this paper preclude adequate description of experimental methods; however, a full report of this research will be ready for publication soon.

### Results and Discussion

Figure 1 shows a typical example of EMG activity during the first sentence of the rainbow passage. Medial pterygoid was the most active muscle, not only in this example, but for all four subjects in almost all speech tasks. This example is typical as medial pterygoid tends to be: 1) moderately active throughout the utterance, 2) most active in relation to jaw closing, and 3) reduced during jaw opening. In Figure 1 masseter and temporalis were slightly active, but in many instances they were quiet throughout the speech sample. When masseter and temporalis were active, it tended to be during jaw-closing movements of large displacement or velocity.

Even though one might hope medial pterygoid activity would increase when the jaw moves further or faster, this is not necessarily the case due to the nonlinear relations between EMG and muscle force (Bigland and Lippold, 1954), and the difficulty in relating jaw-closing forces to parameters of jaw movement. As illustrated for isolated VCs by one subject in Figure 2, medial pterygoid EMG did not increase as a function of displacement. In fact, for the four subjects half the correlations between peak medial pterygoid EMG and displacement (0.15, -0.25, -0.37, and -0.17) and peak velocity (0.53, 0.23, -0.32, and 0.14) were negative. If one assumes that more muscle force is required to increase jaw-closing displacement or velocity, then either this is not reflected in our EMG measurements or is produced by muscles other than medial pterygoid. Figure 2 also shows that for isolated VCs temporalis became more active for larger displacements ( $r = 0.66$ ). Three of the four subjects tended to increase

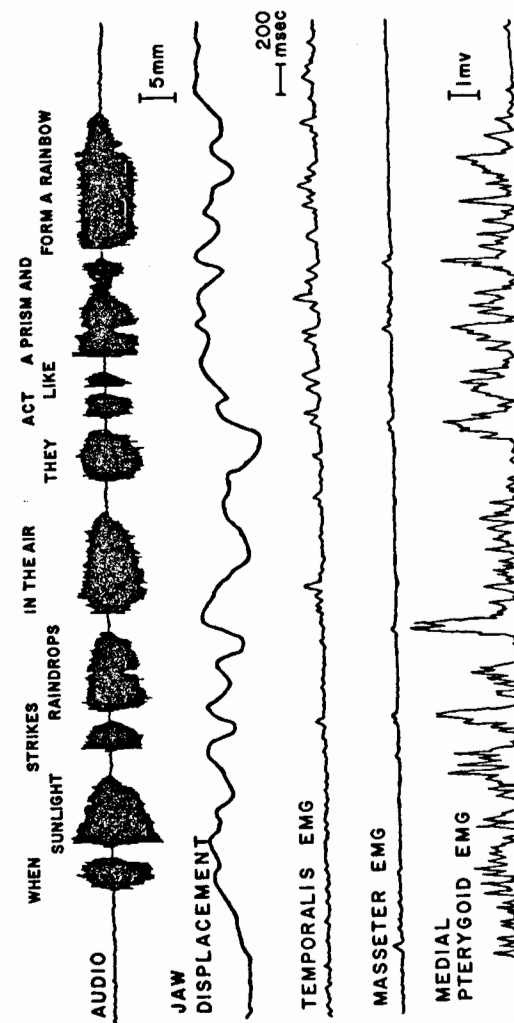


Figure 1. Rectified and smoothed (20 msec TC) EMG during reading for subject 3.



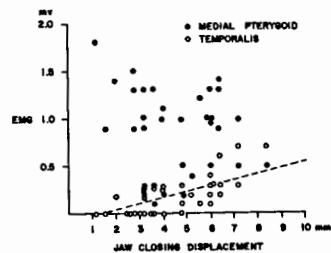


Figure 2. Scatterplot of peak EMG and jaw-closing displacement for isolated VCs by subject 4. The dashed line is a linear regression through the temporalis points. Masseter was zero for all syllables.

temporalis activity for larger (or faster) jaw-closing movements. The other subject tended to use masseter rather than temporalis.

A notable aspect of Figure 2 is the spread in EMG values for movements with similar displacements. Variability is especially evident in repeated syllables with matched displacement and velocity. For example, the subject in Figure 2 repeated [pæ] 24 times in rapid succession. Closing displacement was consistent as one standard deviation was only 16% of the average displacement. One standard deviation of peak velocity was only 14% of average. Mean medial pterygoid activity was 1.05 mv; however, it showed a standard deviation of 43%. Masseter averaged 2.15 mv with a standard deviation of 81%. This is surprising as masseter was quiet throughout the isolated syllables for this subject even though the repeated syllables were well within the range of displacements and velocities for isolated syllables. Variability was evident for most situations, but occasionally subjects were consistent. For example, on the left of Figure 3 one standard deviation of peak medial pterygoid EMG is only 10% of the mean.

In summary, medial pterygoid is consistently more active than masseter and temporalis. However, within this general distinction there appears to be a large amount of utterance-to-utterance variability in the way these muscles are employed.

## Experiment Two

### Introduction

A number of papers have illustrated the abilities of the speech motor control system to compensate for mechanical modifications in movement of the jaw (Folkens and Abbs, 1975; Lindblom, Lubker, and Gay, in press). Both Lindblom et al. and Perkell (1979) suggest that speech motor systems produce appropriate gestures in spite of perturbing factors by employing central stimulation strategies. For example, when one speaks with a bite block, a central movement plan adjusts the roles of many articulators for the lack of jaw movement. As the jaw is fixed with the bite block one might also expect the central movement plan to eliminate "unnecessary" jaw-closing muscle activity. The purpose of this experiment was to record EMG from the jaw-closing muscles with a bite block in place and see if there is a reorganization of muscle activity.

### Method

This experiment was carried out in the same experimental sessions, with the same electrode placements as experiment one. After producing the speech sample with the jaw free to move, the sample was repeated with the jaw fixed with a bite block. Bite blocks providing both 5 mm and 15 mm of interincisor distance were employed.

### Results and Discussion

With both sizes of bite blocks there were consistent bursts of EMG activity which related closely to the temporal patterns of EMG found in each muscle during the jaw free condition. This is illustrated in Figure 3 for medial pterygoid as [pæ] was repeated at a fast rate.

As the mandible is not moving, it is not clear why the phasic jaw muscle activity persists with the bite block. A complete central reorganization would be expected to remove unnecessary muscle activity. As an alternative, it may be that the phasic jaw muscle activity is involved in the organization of other articulatory movements occurring with the bite blocks. That is, peripheral motor control mechanisms (including brainstem reflexes; McClean, Folkens, and Larson, in press) may be important

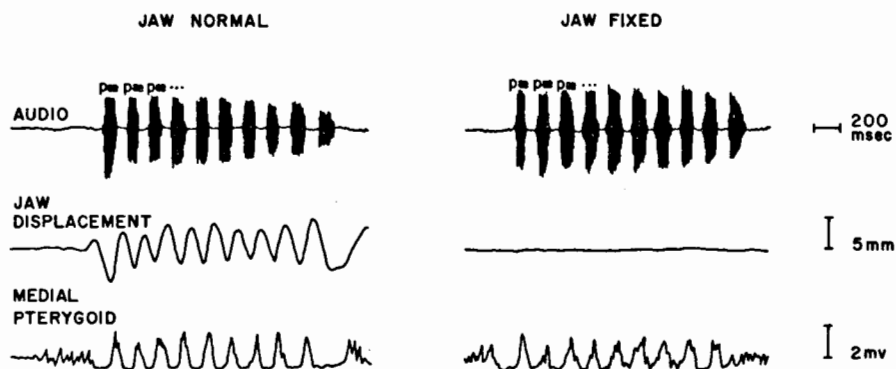


Figure 3. Rectified and smoothed (20 msec TC) EMG for [pæ] repeated at a fast rate by subject 1.

components in the processes which accomplish appropriate speech movements with and without mechanical interferences.

#### References

- Bernstein, N. (1967): The Coordination and Regulation of Movements, New York: Pergamon Press.
- Bigland, B. and O. Lippold (1954): The relation between force, velocity, and integrated electrical activity in human muscles, J. Physiol. 123, 214-224.
- Folkins, J. and J. Abbs (1975): Lip and jaw motor control during speech: Responses to resistive loading of the jaw, JSHR 18, 207-220.
- Hylander, W. (1975): The human mandible: Lever or link?, Am. J. Phys. Anthrop. 43, 227-242.
- Kawamura, Y. (1974): Physiology of Mastication, Basel: S. Karger.
- Lindblom, B., J. Lubker, and T. Gay (in press): Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation, J. Phonetics.
- MacNeilage, P. (1970): Motor control of serial ordering of speech, Psych. Rev. 77, 182-196.
- McClellan, M., J. Folkins, and C. Larson (in press): The role of the perioral reflex in lip motor control for speech, Brain and Language.

Palmer, J. (1972): Anatomy for Speech and Hearing, 2nd Ed., New York: Harper & Row.

Perkell, J. (1979): Phonetic features and the physiology of speech production, in Language Production, B. Butterworth (ed.), New York: Academic Press.

Zemlin, W. (1968): Speech and Hearing Science: Anatomy and Physiology, Englewood Cliffs, New Jersey: Prentice Hall.

EFFECTS OF EFFERENT AND AFFERENT INTERFERENCE ON SPEECH PRODUCTION:  
IMPLICATIONS FOR A GENERATIVE THEORY OF SPEECH MOTOR CONTROL

Thomas Gay and Michael Turvey, Haskins Laboratories, New Haven, Connecticut, U.S.A.

One might claim that speech production proceeds in open-loop fashion: for a given speech sound a motor program prescribes a standard set of instructions to the musculature. Against this claim, however, is the fact that the backdrop of articulatory states into which the standard instructions would be inserted is itself not standardized. The initial conditions (or contexts) for the articulatory gestures yielding a given speech sound vary considerably (cf. MacNeilage, 1970). This is a most notable feature of speakers: within reasonable limits they are capable of producing the necessary configurations of articulatory maneuvers for the sounds of speech even though the departure points for those configurations are ever-varying. Moreover, it appears that the configuration of gestures underlying a desired speech sound can be generated with virtually no experimentation and without the benefit of auditory monitoring. Lindblom and his co-workers (Lindblom et al., 1978) have shown that speakers fitted with bite blocks can produce isolated vowels within the range of variability for normal vowel production, and that satisfactory formant matching occurs on the first pitch pulse of the first attempt.

Clearly, the adaptive, generative nature of articulation is not captured by the notion of open-loop control. Consequently, speech investigators have turned to the claim that the control of speech is closed-loop. In closed-loop explanations, a sensory referent is proposed that relates either to the environmental goal of the articulatory gestures, such as a spatial target or an acoustic pattern, or to the movement-producing commands. (The interpretation of sensory referent as a spatial target is currently the more popular interpretation.) The comparison of sensory feedback with the sensory referent yields an error signal that provides the basis for adjusting the lower level motor mechanism(s) responsible for controlling the referent. Over successive comparisons an increasingly closer match between the feedback sensory signal and the desired sensory referent is achieved.

While a closed-loop mechanism can, in principle, adjust motor instructions to variable initial conditions to attain the referent, it is not immediately obvious how a feedback mechanism that gradually approaches a desired result could underly the immediate adjustment to context evidenced in everyday speaking and underlined by the phenomenon reported by Lindblom and his colleagues. What is needed is a mechanism that: (1) can produce the appropriate articulatory gestures in the face of variable and often novel initial conditions, and (2) can do so without trial and error.

On first thought, these two criteria are met by model-referenced control. Here, the closed-loop mechanism tied to the peripheral speech apparatus is modeled centrally so that motor commands and their sensory consequences can be simulated for the current conditions of the peripheral speech apparatus. The simulated motor commands that result in a match between the simulated sensory feedback and the sensory referent are then realized as actual motor commands. In principle, the predictive simulation of model-referenced control could underly the immediate readjustment phenomenon (Lindblom et al., 1978). There is, however, a potentially serious drawback to any closed-loop explanation: While an error signal can index how near the collective action of a number of muscles is to the desired consequence, it does not prescribe in any straightforward way how the individual muscles are to be adjusted to give a closer approximation to the referent (Fowler and Turvey, in press).

There is another mechanism, very different from closed-loop control, that meets the two criteria noted above. The rationalization and evidence for this mechanism - referred to as a coordinative structure - has been presented elsewhere in some detail (Fowler, 1977; Fowler, Rubin, Remez and Turvey, in press; Turvey, Shaw and Mace, in press). A rough sketch must suffice for current purposes.

Consider a set of several (relatively) independent muscles. As an aggregate, the muscles would exhibit a large number of degrees of freedom and would rely on a source external to themselves for their control. The number of degrees of freedom can effectively be reduced by functionally linking the muscles so that they mutually determine one another's states in a systematic fashion. But such linkage control would, in large part, be internal to the set of muscles. Such functional linkages, that render

an aggregate of relatively independent muscles into a single autonomous unit, may be conceived of as equations-of-constraint written, as it were, on the ascending and descending neural pathways.

To identify some important features of this latter system, let us compare it with closed-loop control in relation to the problem of uttering a vowel under conditions of efferent and afferent interference. In the closed-loop perspective, to produce a given vowel is to specify a particular spatial target as referent. In the coordinative structure perspective just outlined, to produce vowels is to organize the articulators into a single, autonomous system according to a particular equation (or set of equations) of constraint; and, to produce a given vowel is (perhaps) to parameterize that system in a particular way (cf. Fowler, 1977).

Suppose that a speaker impeded by a bite block is requested to utter a given vowel. The model-referenced version of closed-loop control assumes that the condition of the speech apparatus is sensed and motor commands together with sensory feedback are simulated to determine what needs to be done given these conditions. The coordinative structure perspective simply notes that if some parts of the system are 'frozen' the other parts will, by virtue of the equation(s) of constraint, automatically assume values tailored to that of the frozen part and appropriate for producing the vowel.

Suppose now a speaker is interfered with not by a bite block but by anesthetization of parts of the speech apparatus and, as before, is requested to utter a given vowel. In this situation model-referenced control must suffer to the extent that sensory information about initial conditions is not available. In short, anesthetization should impair vowel production considerably more than a bite block restriction. From the coordinative structure perspective, however, anesthetization and bite block should be equivalent in that neither one alone should seriously perturb vowel production. For some members of a coordinative structure to be 'uninformed' about the states of other members is not important; as long as all members of the structure can vary, equilibration according to the equation(s) of constraint will occur and vowel production will be successful. However, we suspect that if some members cannot vary (due to a bite block) and their values are not

communicated within the system (due to anesthetization), then fulfilling the equation(s) of constraint will not be possible and successful vowel production would be seriously hindered. The experiment that follows is a preliminary appraisal of these notions. In it, both efferent and afferent variables were either interfered with directly or controlled indirectly during the production of several isolated vowels. Both acoustic and electromyographic measures were used to determine how speech performance is affected when the linkage among these variables is both partially and completely disrupted.

#### Method

Subjects were two adult male native speakers of American English, one phonetically trained (WE) and the other phonetically naive (SJ). The speech material consisted of the isolated vowels, /i,a,u/. Four separate articulatory variables were controlled directly and one was controlled indirectly. A bite block and an artificial palate were used to produce direct efferent interference, and anesthesia of the temporomandibular joint (TMJ) and oral mucosa were used to produce direct afferent interference. Two different bite blocks were used, one 23 mm long and the other 3 mm long. The longer bite block was used to fix jaw position for the close vowels /i/ and /u/, and the shorter bite block was used for the open vowel /a/. An acrylic artificial palate was constructed from upper mouth casts of both subjects. This prosthesis was approximately 10 mm thick at the midline, 3 mm thick along its edge, and 5 cc in volume. It extended from the posterior surface of the central incisors to approximately 8 mm anterior to the soft palate. Jaw position afference from the mechanoreceptors of the temporomandibular joint was eliminated by 2 ml of xylocaine injected directly into the joint capsules, bilaterally. Oral mucosa sensation was eliminated by spraying the entire oral cavity with a benzocaine solution.

The recording procedures were as follows: First, the subjects produced three triads of each vowel spontaneously, with the bite block, the artificial palate, and the bite block and artificial palate, in that order. Anesthesia was then applied in two steps. For one subject (WE), the joint was anesthetized first, while for the other subject the topical anesthesia was applied first. In each case, the entire vowel sequence was repeated after

each anesthetization. The experiments were run with the subjects seated in front of a microphone. For one subject (WE), electromyographic recordings from the genioglossus (tongue) and orbicularis oris (lip) muscles were obtained using conventional hooked wire techniques. All data were recorded on magnetic tape for later analysis.

### Results

For both subjects, the effects of the experimental conditions were variable and evident only for /i/ and /u/; the formant frequencies of /a/ were virtually unaffected by either mechanical interference or anesthesia. Apparently, only pharyngeal cavity variables are relevant to /a/. For both /i/ and /u/, articulatory performance was unaffected by anesthesia alone, the artificial palate under all conditions of anesthesia, and the bite block under normal conditions and under incomplete anesthesia. Performance was affected, however, and dramatically so, when the bite block was introduced either alone or in combination with the artificial palate under complete anesthesia. These effects were substantial not only perceptually, but at the acoustic and muscle activity level as well. For example, first and second formant frequencies of the first spontaneous /i/ produced by subject WE were 275 and 2275 Hz. These values were approached for all experimental combinations except the TMJ + topical anesthesia + bite block and TMJ + topical anesthesia + bite block + artificial palate conditions where first and second formant frequencies shifted to 375 and 2050 Hz and 425 and 1600 Hz, respectively. Formant shifts were also evident for subject SJ, although to a slightly lesser degree. The EMG data dramatically illustrate these effects. Figure 1 shows the genioglossus EMG for the spontaneous bite block and TMJ + topical anesthesia + bite block conditions for the vowel /i/. The top trace shows the genioglossus muscle activity for /i/ produced spontaneously. The middle trace shows the corresponding EMG for the simple bite block condition. The increase in activity here is expected because the tongue has farther to move from a fixed-open position toward its target. Note also that the increase in activity is present at onset, before any online feedback mechanism would have time to generate an adjusted movement. The lower record corresponds to the TMJ + topical anesthesia + bite block condition. It shows virtually no activity. Absence of muscle

activity was the rule for all tokens within this condition as well as for the TMJ + topical anesthesia + bite block + artificial palate condition. Apparently, fixation of both the efferent and afferent variables resulted in an inability to produce any coordinated movement; hence, a neutral tongue position and a tendency toward schwa.

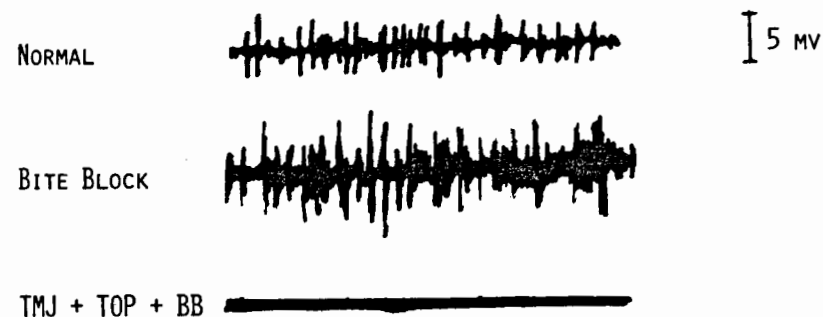


Figure 1

EMG activity for three experimental conditions.

This motor disorganization, however, was relatively short-lived. In each case learning took place and the normal vowel targets were reached after several trials. Table 1 shows the formant frequency values for /i/ produced under the most extreme experimental condition for each of nine token repetitions, for subject WE. Again, measurements were made at the time of the first glottal pulse. For even this extreme condition, complete acoustic compensation was attained by the sixth trial where vowel targets approached those of the spontaneously produced vowel.

Table 1

First and second formant frequencies for the vowel /i/ produced by subject WE under the most extreme experimental conditions (TMJ + topical anesthesia + bite block + artificial palate). Normal vowel target values are: F1 = 275 Hz, F2 = 2275 Hz.

	TRIALS								
	1	2	3	4	5	6	7	8	9
F1	425	500	475	325	325	275	300	300	275
F2	1600	1700	1900	2050	2150	2175	2225	2225	2250

Conclusions

The main finding of this experiment was that interference with either an efferent or afferent variable alone did not affect the production of isolated vowels; however, simultaneous interference with both efferent and afferent variables seriously altered vowel production. It is our view that these findings demonstrate both the necessity of a generative approach to speech production modeling and the utility of a coordinative structure mechanism for the control of speech movements. First, the experimental conditions produced novel physical and sensory situations that were met with immediate and successful articulatory responses. An open-loop model based on stored experiences cannot explain the success of these responses. Second, from the coordinative structure perspective, the finding that afferent interference does not affect vowel production unless an efferent variable is frozen is consistent with the view that these muscles are functionally linked across efference and afference in such a way that control can be taken over by either system when the other is fixed.

References

- Fowler, C.A. (1977): "Timing control in speech production", Indiana University Linguistics Club, Bloomington, Indiana.
- Fowler, C.A. and M.T. Turvey (1978): "Skill acquisition: An event approach with special reference to searching for the optimum of a function of several variables", to appear in Information Processing in Motor Control and Learning, G. Stelmach (ed.), New York: Academic Press (in press).
- Fowler, C.A., P. Rubin, R.E. Remez and M.T. Turvey (1978): "Implications for speech production of a general theory of action", Language Production, B. Butterworth (ed.), New York: Academic Press.
- Lindblom, B., J. Lubker and T. Gay (1978): "Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation", JPh (in press).
- MacNeilage, P.F. (1970): "Motor control of serial ordering of speech", Psychological Review 77, 182-196.
- Turvey, M.T., R. Shaw and W. Mace (1978): "Issues in the theory of action: Degrees of freedom, coordinative structures and coalitions", in Attention and Performance VII, ed. by J. Requin, Hillsdale, N.J.: Erlbaum.

A CORRELATION ANALYSIS OF EMG ACTIVITY AND THE MOVEMENT OF  
SELECTED SPEECH ORGANS

Hajime Hirose, Research Institute of Logopedics and Phoniatics,  
Faculty of Medicine, University of Tokyo, Tokyo, Japan

Introduction

In the study of the dynamic aspects of speech production, it is ultimately necessary to investigate the pattern of motor control signals from the central nervous system and the dynamic characteristics of the speech organ(s) which act(s) in response to the control signals. Although the pattern of the motor control signal has usually been observed in the form of electromyographic (EMG) potentials, the quantitative analysis of the relationship between EMG activity and articulatory movement has remained difficult. Cinefluorographic observation combined with simultaneous recording of EMG signals has been considered to be most satisfactory, but the acquisition of necessary information is generally restricted due to the dosage problem.

The introduction of the x-ray microbeam system to speech research (Kiritani, Itoh and Fujimura, 1975) solved the dosage problem to a large extent and, at the same time, proved useful for reducing the time required for data analysis. Figure 1 shows a data collection and analysis system in use at the Research Institute of Logopedics and Phoniatics, Faculty of Medicine, University of Tokyo.

The present study is an attempt to analyze the dynamic characteristics of the movements of selected speech organs recorded by means of our x-ray microbeam system simultaneously with EMG recordings of the activity of the related articulatory muscles during speech. In the present paper, the preliminary results of an analysis of velar movement will be presented as an example, in reference to the pattern of the EMG activity of the levator palatini muscle. The articulatory movement of the velum is known to be controlled almost solely by the activation and suppression of the levator palatini. Velar movement is relatively independent of the movement of the other articulators and, thus, relationship between the displacement of the velum and the EMG patterns of the levator palatini can be considered to be relatively straightforward.

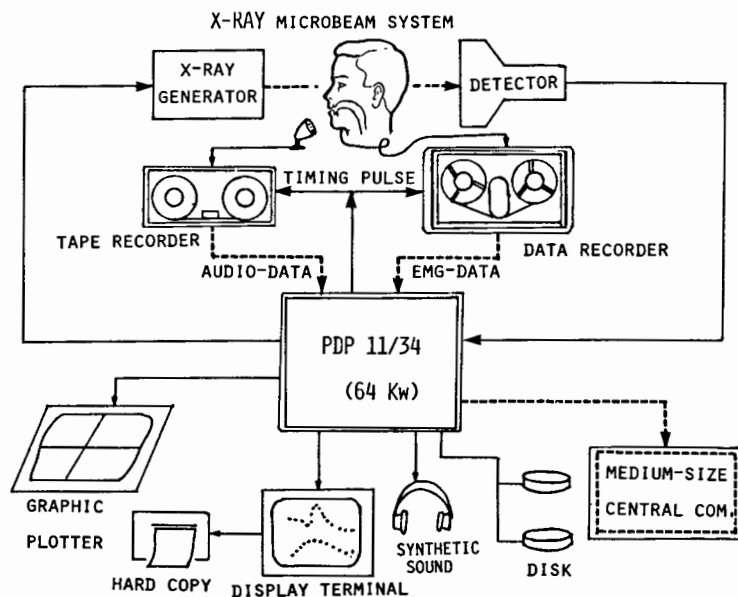


Figure 1 The Articulatory data collection and analysis system at the RILP, University of Tokyo

#### Data recording

An adult male speaker of the Tokyo dialect served as the subject of the present study. The subject read a list of test words either in isolation or embedded in appropriate frame sentences.

For recording the articulatory movements, the movement of lead pellets attached to the pertinent articulators was tracked and recorded by means of the computer-controlled x-ray microbeam system. For recording velar movement, a strip of thin plastic film with a lead pellet attached to its end was passed through a nostril, placing the pellet on the nasal surface of the velum. The pellet movement was recorded with a frame rate of approximately 190 frames/sec.

Conventional hooked-wire electrodes were inserted into the levator palatini in this particular case. The EMG signals were recorded on an FM data recorder together with the speech signals

and the timing pulses which were generated from the computer for each frame of the x-ray tracking. The EMG signals were then digitized through an A/D converter with a sampling rate of 8 kHz. Absolute values were taken and integrated over 5.83 msec, the value of which corresponds to the interval between successive timing pulses.

#### Data analysis

The Y-coordinate of the pellet on the velum was selected to represent the movement of the velum, and the relationship between the time function of the coordinate value and the EMG signal was examined. EMG activity is associated with the generation of muscle force and, therefore, can be related to such variables as the displacement, velocity and acceleration of the movement of the pertinent articulator. Thus, the present analysis aimed at obtaining a quantitative estimation of the relationship between these variables and the EMG signal.

It was assumed that the EMG activity of the levator palatini at a given instant could be expressed as the sum of the three components dependent on the displacement ( $y$ ), velocity ( $\dot{y}$ ) and acceleration ( $\ddot{y}$ ) of the movement of the velum. Thus, an estimated EMG signal at a given time can be given as in equation (1).

$$\hat{E}_i = c_0 + c_1 y_i + c_2 \dot{y}_i + c_3 \ddot{y}_i \quad \text{--- (1)}$$

In this equation, the subscript  $i$  denotes the  $i$ -th time sample. The above equation indicates that velar movement is realized as the response of a linear second order system to the EMG signal which is given as input. The coefficients which give the best approximation were estimated by the least square error method. That is, for every time sample of EMG signal  $E_i$ , the estimate  $\hat{E}_i$  in equation (1) was formed by using the coordinate values obtained by x-ray tracking. The coefficients,  $c_{0-3}$  were determined by minimizing the value of error ( $E_{rr}$ ) in equation (2).

$$E_{rr} = \sum_i (E_i - \hat{E}_i)^2 \quad \text{----- (2)}$$

In the above procedure, it was necessary to introduce a temporal smoothing of the observed coordinate value, since, without smoothing, the noise components in the calculated velocity and



acceleration were so dominant that virtually no effective correlation could be observed between these variables and the EMG signals. In order to reduce the noise effect, the temporal variation of the coordinate value within a short time window was approximated by the parabolic function of time. In the data sets obtained in the present study, it was found that the error was minimum for a time window of about 30 frames. Thus, in the present analysis, the values calculated for this time window width were considered to be the best estimates of the coefficients.

#### Results and discussion

The characteristic constants of the linear second order system were calculated from the estimated values of the coefficients in equation (1). The value of the damping factor was found to be close to 1, which implied that the second order system is critically damped. The characteristic time constant was approximately 80 msec, regardless of the difference in speaking rate.

Figure 2 shows examples of the x-ray and EMG data obtained. These curves correspond to the three different types of test words, /bemeē/, /beN'ee/ and /beNmee/, each of which was embedded in a frame "sorewa \_\_\_\_\_ desu" (that is \_\_\_\_\_). In the test words, /N/ represents the syllable final nasal element in Japanese. The sequence of nasal segments /Nm/ is generally uttered as a geminate nasal consonant. For each test word, the curve at the top shows the audio signal, the second and third curves are the temporal patterns of the Y-coordinate value for the lower lip and the velum, respectively. The bottom curve shows the integrated raw EMG, with the estimated EMG curve calculated by using equation (1) superimposed.

It can be seen that in /bemeē/ the velum lowering for /m/ starts immediately after the oral release of the initial /b/ and continues until the release of /m/. Velar elevation then begins, the speed of which appears to be slower than that of lowering, and, as a result, the temporal pattern of velar movement is asymmetrical for the production of /m/. In /beN'ee/, the velum lowering for /N/ continues longer, and the velar displacement is larger than for /m/ so that the temporal pattern appears to be symmetrical. In /beNmee/, the level of the maximum velum lowering for /Nm/ is higher than for /N/ in /beN'ee/, although the duration of nasaliza-

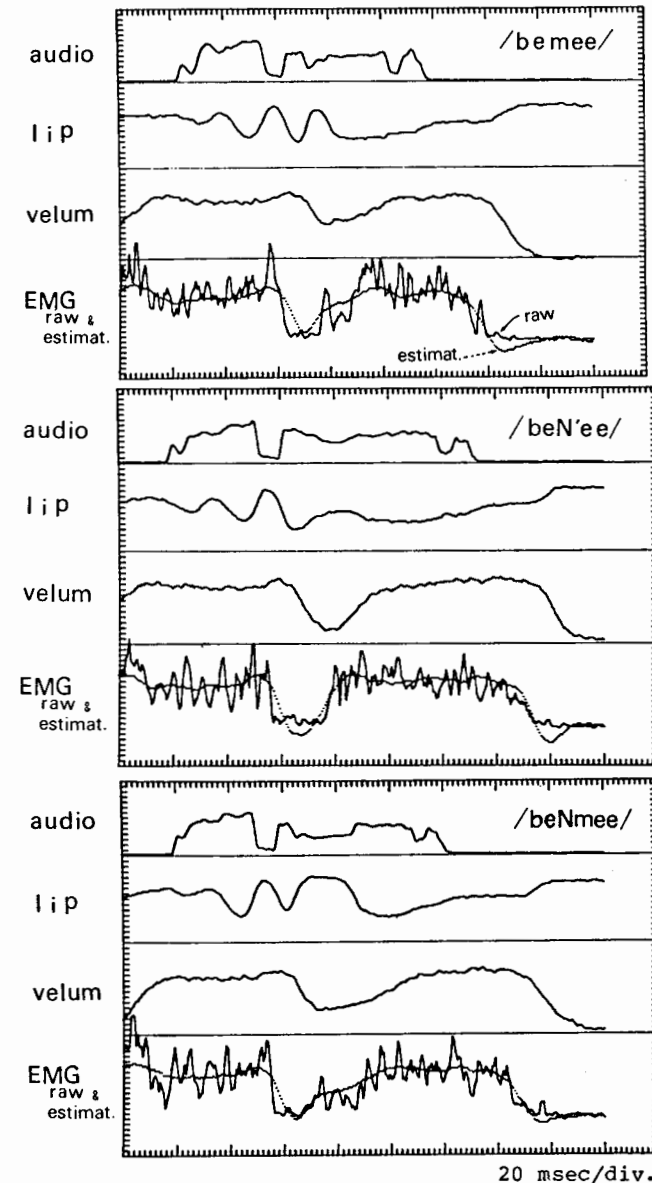


Figure 2 The temporal patterns of the Y-coordinate value for the lower lip and the velum, compared with the audio signal (top) and the raw and the estimated EMG curves (bottom).

tion is longer. In this case, after reaching the level of maximum lowering, the velum appears to stay at, or to ascend very gradually from, that level and, thereafter, it ascends with a speed similar to that for /m/ in /bemeē/.

Comparing the patterns of the raw EMG activity with those of the velar movement, it appears that the pattern of levator activity for /N/ in /beN'ee/ is characterized by a step-like EMG suppression, the level of which is the same as that for the resting position of the velum after the cessation of the utterance. In other words, the movement of the velum for /N/ can be regarded as a smoothed response of the velum as a second order system to the step-like control signal to the velum. In /beNmee/, on the other hand, a rapid suppression of levator activity is followed by a short period of an intermediate level of EMG activity. Although the EMG level of the initial part of the suppression for /Nm/ is apparently the same as that for /N/ in /beN'ee/, the velum does not show an extreme lowering but stays at a somewhat higher position. Thus, it appears that the period of intermediate level EMG activity mentioned above is responsible for the characteristic pattern of velum movement for /Nm/. The duration of the suppression of the levator activity for /m/ in /bemeē/ is relatively short compared to the estimated value of the time constant of the second order system, and, therefore, the pattern of the velar movement for /m/ can be taken as a ballistic impulse response to the EMG signal. However, as seen in Figure 2, a short re-activation of the levator palatini reaching the intermediate level is observed in some utterance samples of /bemeē/. Thus, for /m/, a question still remains as to whether the apparent re-activation of the levator palatini may necessarily result in the gradual ascent of the velum after the initial lowering.

The characteristic pattern of each of the three utterance types is also demonstrated in the estimated EMG curve. In particular, the estimated curve for /Nm/ is characterized by the fact that when the EMG activity increases after the negative peak of suppression, the rate of increase is temporarily depressed before it reaches the maximum level, and, as a result, there is a plateau in the estimated curve. This result would indicate that, as far as the second order linear relationship between the levator EMG and velar movement is concerned, the patterns of the velar movement

for /Nm/ can not be accounted for by a constant increase in the EMG activity after suppression. Rather, it can be assumed that an intermediate stage of EMG control is necessary during the phase of re-activation.

It has been reported that there are characteristic differences in the temporal patterns of velar movements for these three utterance types (Ushijima and Hirose, 1974; Fujimura, Miller and Kiritani, 1976). The result of the present study seems to confirm this result. It also suggests that these differences are based on the different patterns of motor control to the velum.

For a better understanding of the dynamic aspects of speech production, further attempts should be made to accomplish the quantitative analysis of the relationship between EMG signals and articulatory movements. A preliminary analysis of jaw movements in reference to the pattern of the EMG activity of the related muscles is now in progress.

#### References

- Fujimura, O., J.E. Miller and S. Kiritani (1976); "Syllable final nasal element in English - An x-ray microbeam study of velum height", 92nd Meeting of ASA, S 65.
- Kiritani, S., K. Itoh and O. Fujimura (1975); "Tongue-pellet tracking by a computer-controlled x-ray microbeam system", JASA 57, 1516-1520.
- Ushijima, T. and H. Hirose (1974); "Electromyographic study of the velum during speech". JPh 2, 315-326.

#### Acknowledgement

This research was supported in part by Grant in Aid for Scientific Research, Ministry of Education (No.349008).

ON THE USE OF OROSENSORY FEEDBACK: AN INTERPRETATION OF  
COMPENSATORY ARTICULATION EXPERIMENTS

Joseph S. Perkell, Research Laboratory of Electronics,  
Massachusetts Institute of Technology, Cambridge, Mass. 02139,  
U.S.A.

A number of experiments have been performed on "compensatory articulation" with the aim of understanding more about speech motor programming. Several of these experiments have used bite blocks to constrain the mandible in abnormally open (or closed) positions while the subjects produced steady state vowels (/i/, /a/, and /u/) (cf. Lindblom, 1971; Lindblom, et al., 1977; Lindblom et al., in press; Gay and Turvey, these proceedings). The resulting formant patterns were measured at the first glottal pulse to avoid any possible effects of auditory feedback (which was not masked out). It was found that vowels produced with significantly abnormal jaw openings (i.e. 22-25 mm open for /i/) were essentially the same in quality as those produced normally by the same subjects. However when bite blocks were used in conjunction with oral topical anesthesia (Lindblom, et al., 1977) or with a combination of oral topical anesthesia and anesthesia of the temporomandibular joint (Gay and Turvey, these proceedings), subjects needed several attempts to produce appropriate vocal tract configurations and sound outputs. In the latter experiment, the application of oral topical anesthesia alone was not enough to impair subjects' ability to produce vowels appropriately.

Lindblom and his co-workers interpret their findings as support for the following view of the role of orosensory feedback. Tactile information from the labial and oral mucosa can be utilized in the motor programming of speech. Vowel "targets" may be encoded as [oro]sensory goals which reflect a neuro-physiological encoding of area functions. These goals serve as a basis for the elaboration of motor commands by structures which "can generate appropriately revised motor commands on the basis of the feedback positional information available before onset of phonation" (Lindblom, et al., in press).

These results and their interpretations must be viewed with caution for a number of reasons. For example, a generous application of topical anesthesia to the oral and pharyngeal cavities can have a distracting effect on the subject (Lindblom,

personal communication). Perhaps more importantly, a steady-state paradigm which allows the subject time to "organize" his response before presenting it may reflect functions which are not part of normal dynamic speech motor processes (cf. Leanderson and Persson, 1972; Abbs and Eilenberg, 1976). Nevertheless, the results are provocative enough to warrant further examination, particularly in light of a recent experiment on arm movements and another experiment on compensatory articulation.

Polit and Bizzi (1978) have performed an experiment in which 3 adult monkeys were trained to point to a target light with the forearm and hold the position for about 1 second in order to obtain a reward. The monkey could not see its forearm which was fixed to an apparatus that permitted only flexion and extension about the elbow in the horizontal plane. Performance was tested before and after a dorsal root section which eliminated somatosensory feedback from both upper limbs. In both intact and deafferented animals, the arm was unexpectedly displaced within the reaction time of the monkey, and in both cases the displacement of the initial arm position did not affect the attainment of the intended final steady-state position. These results suggested to the authors that a central program specified an equilibrium point corresponding to the interaction of agonist and antagonist muscles. A change in the equilibrium point leads to movement and attainment of a new posture.<sup>1</sup> However, it was also found that when the spatial relationship between the animal's arm and body was changed, the pointing response of the deafferented monkeys was inaccurate, and remained so even when visual feedback was allowed. In contrast, the intact monkeys were able to compensate within a few tries to the new position without visual feedback. This finding suggested that one major function of afferent feedback is in the adaptive modification of learned motor programs (Polit and Bizzi, 1978).

Following these authors' interpretation of their results, we might consider that in establishing the central program for the performance of the motor task (i.e., learning the task), the monkeys were incorporating a subconscious "knowledge" of the relationships between the target points with respect to the

(1) The existence of additional processes related to the dynamic aspects of movements is acknowledged, but not treated by Polit and Bizzi (1978).

apparatus and the muscular settings which would result in correct pointing. In doing so, the monkeys were calibrating the biomechanical properties of the system with respect to a particular frame of reference (i.e. orientation in space in relation to the body) with the use of somatosensory feedback from the system. When the frame of reference changed, only the monkeys with intact somatosensory pathways were able to "recalibrate" the central program to the new frame of reference.

This line of reasoning and the interpretations of Lindblom and his colleagues lead us to a possible, slightly more specific explanation of the compensatory articulation results. In the case of steady-state vowel productions, the frame of reference is defined as the configuration of the dorsal walls of the vocal tract and the position of the mandible. The target (or goal) consists of a vocal-tract area function as sensed by a complex pattern of sensory feedback from the vocal tract. Normally, to produce a steady-state configuration, the control mechanism has a choice of: 1) using a pattern of peripheral feedback to compare with one that has been learned in association with a particular area function and vowel quality, or 2) using a set of equilibrium levels of muscle excitations. These muscle excitation levels can be stored or computed on the basis of an overlearned knowledge of the vocal-tract geometry and biomechanical properties.

Now let us consider the three possible combinations of the use of anesthesia and/or bite blocks. With only (complete) anesthesia, the controller uses option 2. In other words, with a frame of reference which is assumed to be normal, the controller is still capable of specifying equilibrium muscle excitations which it "knows" will produce the correct area function. On the other hand with only the bite block, the controller uses option 1. The appropriate area function is produced by comparing peripheral feedback with the "known" pattern. With anesthesia and the bite block, neither option is available. The frame of reference has been changed. The absence of feedback about the new frame of reference precludes an a priori recomputation of appropriate equilibrium muscle excitations, and the absence of tactile feedback precludes a direct comparison with the known pattern. This last statement is reinforced by the results of Gay and Turvey in which only combined anesthetization

of the oral mucosa and the temporomandibular joint (along with the bite block) rendered the subject incapable of producing the vowel correctly on the first try. The loss of joint sensation would eliminate the feedback about the frame of reference, needed for a recalibration of the central program, and the loss of sensation from the oral mucosa would preclude using such feedback directly in an error-minimizing feedback loop.<sup>2</sup>

The hypothetical use of afferent information to keep the controller informed about the frame of reference would be equivalent to the function proposed by Polit and Bizzi (1978) in the adaptive modification of learned motor programs. Presumably, the predictive simulation mechanism proposed by Lindblom, et al., (in press) also needs to use feedback in a similar way. It has been suggested by numerous investigators that learning a motor activity consists in part of substituting central programming for the use of peripheral feedback. This use of central patterning presumably incorporates an ability to adjust the parameters of the central program to account for changes in the frame of reference. In the case of speech, such changes correspond to speaking with a pipe clenched between the teeth, with the head tilted to one side, or resting ones' chin in his or her hand.

Gay and Turvey also found that /i/ and /u/ productions are affected by the combination of joint and topical anesthesia with a bite block while /a/ is not. This difference might be explained by their report that the topical anesthesia was applied to the oral cavity, where the acoustically most critical points of maximal constriction for /i/ and /u/ are located. (A given change in the dorsal-ventral location of the tongue surface will have a proportionally larger effect on the vocal-tract cross-sectional area at the point of maximum constriction than at other locations where the cross-sectional area is greater.) If the anesthesia did not exert a strong effect at the point of maximal constriction for /a/, we might expect it to be produced normally, with the use of feedback which is less consciously obvious but still available from that region. The importance of the pattern of contact at

(2) The fact that only topical anesthesia in combination with bite block was sufficient to impair vowel production in the subject of Lindblom, et al. (1977) might be due to individual differences or differences in the extent and depth of topical anesthesia.

the point of maximal constriction for the vowel is further suggested in lateral cineradiographic tracings (Netsell, et al., 1978) of normal and compensatory productions of the vowel /i/ for 3 subjects. For each subject the normal and compensatory dorsal tongue contours show considerable overlap and the overlap is most pronounced at or near the point of maximal constriction.

This brief analysis greatly oversimplifies the issues in a number of respects. It relies on a small amount of data. It overlooks the significant differences between deafferentation and the application of anesthesia as well as the unnatural nature of both experimental paradigms. And as we have mentioned, it deals with a steady-state task which may be quite different from anything actually found in speech. For these reasons, we must be very tentative in extending our interpretations to cover normal articulatory movements. However, on the basis of considering a number of additional aspects of speech production and the control of movement (see Perkell, 1979a), it is possible to offer the following speculations on the use of orosensory feedback.

1) Orosensory feedback may play a role in determining the nature of some distinctive features. It is possible that certain well-defined patterns of orosensory feedback (such as contact of the tongue with maxillary structures) facilitate the production of sounds which have distinctive acoustic and auditory-perceptual correlates (see also, Stevens and Perkell, 1977, Perkell, 1979b). Such patterns of orosensory feedback could be the speech production correlates of distinctive features. Specifications of utterances in the form of feature-related complexes of orosensory goals might serve as a basis for the production of articulatory movements. Thus, orosensory feedback on a long-term basis might be necessary for the establishment and maintenance of a sub-conscious "knowledge" of the orosensory correlates of the features. This "knowledge" could be used directly as suggested by the bite block results or indirectly in the establishment and maintenance of central programs.

2) As suggested by the discussion in this paper, orosensory feedback might be important in informing any central programming mechanism about the overall state of the system or frame of reference. The use of feedback to make adjustments for changes in the frame of reference could cover a time span corresponding

to several movements in a sequence (Polit and Bizzi, 1978). In more general terms, perceptual feedback "regarding the position or movement trajectories of one or more articulators could be used for preprogramming movements several hundred milliseconds into the future." (Larson, personal communication).

3) This paper has implied that central programming plays a major role in the production of articulatory movements. Much of the experimental and theoretical work on other forms of movement suggests that central programming along with internal feedback (feedback entirely internal to the central nervous system) is used for the moment-to-moment (context-dependent) programming of rapid movement sequences. While this is most likely the case for "learned" or "skilled" motor behavior such as speech production (cf. Lindblom, et al., in press), we must keep in mind that vocal tract motor control mechanisms may conceivably have capabilities that other systems do not have (cf. Folkins and Abbs, 1975, 1976, McClean, et al., in press). Thus it is possible that orosensory (peripheral) feedback from the vocal tract is used on a moment-to-moment basis to assist in the programming of articulatory movements in ways that have not been demonstrated for other types of movement.

Ideas such as these are closely related to questions about the nature of fundamental units which underlie the programming of speech production (cf. MacNeilage, 1970). Thus, the difficulty of testing such hypotheses should not stand in the way of exploring them further. The striking similarities between the compensatory articulation and arm movement results discussed above suggest that we may learn increasingly more about speech by continuing to follow future work on analogous types of movement.<sup>3</sup>

#### References

Abbs, J.H. and G.R. Eilenberg (1976): "Peripheral mechanisms of speech motor control", in Contemporary Issues of Experimental Phonetics, 139-168, N.J. Lass (ed.), New York: Academic Press.

---

(3) I am very grateful to Profs. Emilio Bizzi of M.I.T. and Stephanie Shattuck-Hufnagel of Cornell University for their comments on parts of this manuscript. I also thank Dr. Charles Larson of the University of Washington for his very helpful comments on a previous manuscript. This work was supported by National Institutes of Health Grant NS04332.

- Folkins, J.W. and J.H. Abbs (1975): "Lip and jaw motor control during speech: Responses to resistive loading of the jaw", JSHR 18, 1, 207.
- Folkins, J.W. and J.H. Abbs (1976): "Additional observations on responses to resistive loading of the jaw", JSHR 19, 820-821.
- Leanderson, R. and A. Persson (1972): "The effect of trigeminal nerve block on the articulatory EMG activity of facial muscles", Acta-Oto-Laryngologica 74, 271-278.
- Lindblom, B.E.F. (1971): "Neurophysiological representation of speech sounds", Publication 7, Papers from the Institute of Linguistics, University of Stockholm.
- Lindblom, B., R. McAllister, and J. Lubker (1977): "Compensatory articulation and the modeling of normal speech production behavior", Paper presented at the Symposium on Articulatory Modeling, Grenoble, France, July 11-12.
- Lindblom, B., J. Lubker and, T. Gay (in press): "Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation", JPh.
- MacNeilage, P.F. (1970): "The motor control of serial ordering in speech", Psychological Review 77, 182-196.
- McClellan, M.D., J.W. Folkins, and C.W. Larson (in press): "The role of the perioral reflex in lip motor control for speech", Brain and Language.
- Netsell, R., R. Kent, and J. Abbs (1978): "Adjustments of the tongue and lips to fixed jaw positions during speech: a preliminary report", Paper presented at the Conference on Speech Motor Control, University of Wisconsin, Madison, June 2-3.
- Perkell, J.S. (1979a): "Phonetic features and the physiology of speech production", in Language Production, B. Butterworth (ed.), New York: Academic Press.
- Perkell, J.S. (1979b): "On the nature of distinctive features: implications of a preliminary vowel production study", in Frontiers of Speech Communication Research, B. Lindblom and S. Ohman (eds.), London: Academic Press.
- Polit, A. and E. Bizzi (1978): "Processes controlling arm movements in monkeys", Science 201, 1235.
- Stevens, K.N. and J.S. Perkell (1977): "Speech physiology and phonetic features", in Dynamic Aspects of Speech Production, 323-341, M. Sawashima and F.S. Cooper (eds.), University of Tokyo Press.

MOTOR UNIT DISCHARGE PATTERNS DURING SPEECH: TEMPORAL REORGANIZATION DUE TO COARTICULATORY AND PROSODIC EVENTS

Harvey M. Sussman, Department of Linguistics, University of Texas, Austin, Texas, U.S.A.

The functional unit of muscle contraction, and hence movement control, is the motor unit. A motor unit consists of an alpha motoneuron, located, in the case of speech muscles, within a motoneuron pool in the brainstem, and the single muscle fibers innervated by the axonal branches of that motoneuron. For the past few years my colleagues and I have been interested in describing how motor unit discharge properties change to meet the demands of rapid tension development characteristic of speech production. Observation of motor unit discharge activity represents the closest look at the encoding operations of the central nervous system as the peripherally recorded muscle action potentials (MAPs) of motor units (MUs) stand in a 1:1 relationship to discharges of centrally located alpha motoneurons.

Data will be restricted to MU events within the anterior belly of digastric (ABD), a muscle involved in lowering the jaw during speech. Specially constructed intramuscular wire electrodes, designed to facilitate recording at high force levels, were used in all studies. All measurements were performed on a digitized oscilloscopic display (maximum resolution = 0.1 msec) utilizing computer software routines written for temporal and statistical analyses of motor unit/articulator events during ongoing speech (complete details of recording and measurement procedures can be obtained from Sussman et al, 1977).

Recruitment Order

A current view explaining activation of individual MUs is the Size Principle (Henneman et al, 1965). Briefly stated, the



Size Principle holds that MUs are activated according to motoneuron size, with the smallest neurons (having the lowest excitability thresholds) discharging first, followed by successively larger motoneurons. Evidence supporting this view has been extensively gathered, but primarily from animal experimentation. Recruitment order of MUs active during human isotonic movements, such as speech, have not received much attention to date. Our data overwhelmingly supports the Size Principle. Recruitment order of ABD MUs was observed to be fixed and based on size (as determined by peak-to-peak amplitudes of MAPs). The consistency of MU activation for jaw lowering represents an invariant aspect of the encoding program for speech.

Data on recruitment order can also be related to various aspects of articulatory dynamics. Both jaw displacement and velocity were found to be positively related to the number of MUs active (Sussman et al, 1977). In addition, the initial interspike interval (ISI) of the third recruited MU has consistently been shown to be linearly related to both jaw displacement and velocity. During jaw lowering for the initial vowel in /aepae/ tokens, it was found that jaw displacement and velocity increased as the initial ISI decreased. Correlation coefficients ranged from  $-.44$  to  $-.67$  and were significant beyond the  $p < .05$  level for all utterances examined. The relationship between discharge rate of a MU and some aspect of articulatory dynamics was only found for the larger and later recruited MUs (specifically the third MU recruited). The smaller first and second MUs recruited did not exhibit a straightforward relationship between its initial firing rate and jaw movement. Since the larger and later recruited MUs add a proportionately larger contribution to overall tension development (i.e. larger MUs have higher twitch tension levels)

compared to initially active, smaller MUs, it is not surprising to notice movement variables being influenced by discharge characteristics of the larger MUs only.

#### Temporal Reorganization: Coarticulatory Influences

The temporal interval separating activation of the first recruited MU and the initiation of jaw lowering for an open vowel such as /ae/ can be a valuable dependent variable in providing a glimpse at the time program applied to the events of speech motor control. Such an analysis was made for 64 utterances of /aepae/ with separate measurements taken for initial vowel lowering and final vowel lowering. The results are schematically illustrated in Figure 1. For all utterances the first discharge of the first recruited MU occurred approximately 40 msec after the jaw began to lower for V1, and approximately 28 msec prior to jaw lowering for V2 opening gestures. These consistent differences (across three subjects) can be related to the differences in peripheral biomechanics existing at the moment of jaw lowering for the initial preconsonantal vowel versus the final postconsonantal vowel. It is well known that the jaw exhibits anticipatory coarticulation for an open vowel in final position of VCV tokens (Sussman et al, 1973) Thus, the jaw is lowering for the postconsonantal vowel from a position that is considerably lower than the jaw position preceding the initial preconsonantal vowel. Abbs and Eilenberg (1976) have shown that the mechanical advantage of ABD in exerting a lowering force on the mandible decreases the more the jaw is lowered. This reduction of mechanical advantage represents a diminution of the effective muscle force of ADB to bring about additional lowering for the postconsonantal vowel /ae/. The earlier activation of the initially recruited MU for the final postconsonantal vowel as compared to the initial preconsonantal



vowel may reflect a temporal adjustment of the motor time program needed to partially offset the less favorable mechanical advantage of the jaw during this time. It is consistent with this hypothesis that there was a highly significant positive correlation ( $r = .74, p < .01$ ) between jaw position during the medial consonant and temporal onset of MU I. Thus, the lower the jaw immediately prior to subsequent lowering for the postconsonantal vowel, the earlier did MU I activity begin for jaw lowering for V2. This example provides the first illustration of temporal reorganization, on the cellular level, to a behavioral and biomechanical aspect of the encoding program for speech.

Temporal Reorganization: Stress

Previous studies investigating articulatory reorganization due to stress have shown that higher levels of integrated EMG signals, higher rates of articulator movement, and closer approximations of intended target positions accompany high stress conditions. Until recently, there have been no descriptions of temporal change in motor unit discharge patterns due to the prosodic application of syllable stress.

A subject repeated /aepae/ twenty times with equal "moderate" stress on each syllable and twenty times with heavy stress on the second syllable. The first three recruited MUs were examined for both stress conditions. Figure 2 shows recruitment latencies separating the initial discharges of the first three recruited MUs in conjunction with the temporal onset of jaw lowering for the second syllable for both /aepae/ and /aepæe/ tokens. A temporal starting point,  $t = 0$ , was taken to be the onset of MU I's initial spike. For stressed utterances there was a consistent shortening of the intervals separating successively recruited MUs and a shorter latency between MU I's initial discharge

and the moment of jaw lowering for /ae/. Table 1 gives data characterizing the temporal reorganization pattern in terms of means (in msec), standard deviations, and variability coefficients ( $SD/\bar{X}$ ). Not only was the time program advanced for the stressed condition, but, in addition, there was a marked reduction in variability. The percent reduction in variability, calculated by comparing the unstressed and stressed variability coefficients, revealed a 60% reduction for the MU I - MU II interval, a 82% reduction for the MU I - MU III interval, and a 62% reduction for the MU I - jaw lowering interval.

Other changes in discharge characteristics accompanying stress were an increase in mean firing rate (impulses/sec) for each MU and a decrease in the mean initial interspike interval. This later parameter is indicative of a higher instantaneous discharge rate ( $1/\text{initial ISI}$ ) for the stressed syllables. In addition to the higher instantaneous discharge rates for all three MUs, there was also a marked reduction in the variability of the initial ISI, with a progressively larger percent decrease in variability coefficients with recruitment order -- 21% reduction for MU I, 30% for MU II, and 42% for MU III. Thus, there was a "tighter" control over the onset times for the larger and later recruited MU.

These preliminary findings showing a sharp reduction in the variability of recruitment intervals (e.g. MU I-MU II), activation intervals (MU I-jaw lowering), and initial interspike intervals, add a new dimension to our understanding of articulatory movements underlying syllable stress. In addition to the connotations that go along with the familiar Ohman notion of "an instantaneous addition of a quantum of physiological energy" (Ohman, 1967, p. 33) for stressed productions, the encoding program for speech,

as observed in our data, suggests a more precise control of the timing of cellular events, as well as a more forceful execution of the peripheral dynamics.

Motor unit events and their systematic changes during various conditions of ongoing speech can be sensitive indicants of higher level linguistic conditions. Alterations in the temporal program underlying muscle and hence articulator activation can be observed at the level of the alphamotoneuron (at least indirectly that is).

#### References

- Abbs, J. H. and G. R. Eilenberg (1976): "Peripheral mechanisms of speech motor control," in Contemporary issues in experimental phonetics, N. J. Lass (ed.), 139-168, New York: Academic Press.
- Henneman, E., G. Somjen and D. O. Carpenter (1965): "Functional significance of cell size in spinal motoneurons," J. Neurophys. 28, 560-580.
- Ohman, S. (1967): "Word and sentence intonation: A quantitative model," STL-QPSR 2-3, 20-54.
- Sussman, H. M., P. F. MacNeilage, and R. J. Hanson (1973): "Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations," JSHR 16, 397-420.
- Sussman, H. M., P. F. MacNeilage, and R. K. Powers (1977): "Recruitment and discharge patterns of single motor units during speech production," JSHR 20, 613-630.

Table I: Means (in msec), standard deviations, and variability coefficients (SD/ $\bar{X}$ ) for various temporal intervals characterizing motor unit/articulatory events during unstressed and stressed tokens.

	MU I → MU II	MU I → MU III	MU I → Jaw Lowering
Unstressed	$\bar{X}$ 31.9	40.9	53.8
	SD 23.2	41.3	53.3
	VC .7273	1.0098	.9907
Stressed	$\bar{X}$ 23.7	26.6	33.4
	SD 6.8	4.9	12.8
	VC .2869	.1842	.3832

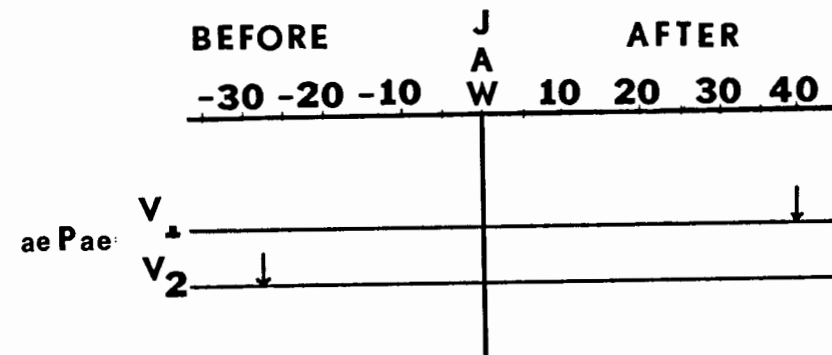


Figure 1: Temporal onset (msec) of initial discharge of the first recruited motor unit with respect to jaw lowering for initial (V1) and final (V2) vowel.

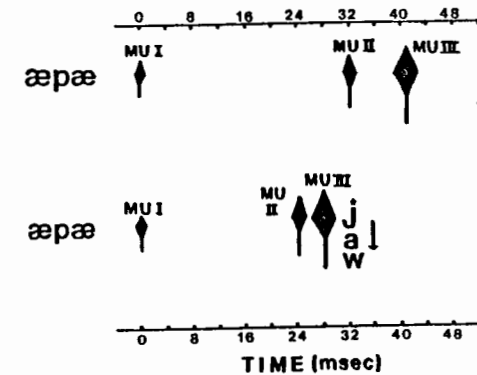


Figure 2: Recruitment latencies separating initial discharges of three recruited motor units during normal and stressed [æpæ]. Asterisks show onset of jaw lowering for open vowel of second syllable.

## THE RELATION BETWEEN SENTENCE PROSODY AND WORD PROSODY

## Summary of Moderator's Introduction

Eva Gårding, Swedish Research Council for Humanistic and Social Sciences and Phonetics Department, Lund University

The contributions to this symposium cover different kinds of prosodic systems and represent different methods of analysis and description. As a moderator I am taken aback by this variety but it may of course help us reach the goal set by the organizers: A pursuit of universal features in the relation between word prosody and sentence prosody. This requires a certain minimum of common terminology for the basic concepts and aspects of prosody. In my summaries and the points suggested for discussion I have used terms from the table below. Synonyms from other authors have been put within parentheses.

Word level	Sentence level	
	local effect	overall (general) effect
tone:	tone (term. junct.):	intonation:
contour tones	downglide	downdrift (declination)
level (static)		absence of downdrift
H (High)	upglide	
L (Low)		
accent (stress):	accent (stress)	
pitch accent		
?		
Citation form (ideal manifestation, ideal shape): a lexical item plus declarative sentence intonation and sentence accent		
Basic form: abstract representation or concrete form freed of other prosodic features		

SummariesThai. A. Abramson

The author is mainly interested in the question of whether the effects of sentence prosody are strong enough to weaken or destroy the lexical tones, the shapes of which have been derived from citation forms. Apart from perturbation from initial consonants there is perturbation from neighbouring forms. Also, in compounds one tone may be replaced by another (sandhi). In running

speech the tones are preserved but their shapes may undergo severe distortions. Sentence intonation can be marked locally by final particles and final sentence tone (terminal junctures), e.g. rising pitch for doubt, falling for finality, sustained for continuation and/or by overall effects. The author examines interspeaker frequency shifts in terminal junctures, expressed as percentages of voice range. Sentence accent is marked by lengthening, increase of amplitude and ideal tone shapes.

Some Southern Nigerian Languages. K. Williamson

The seven languages reported on have two basic tones, H and L. Two have distinctive downstep, H<sup>↓</sup>H versus HH. Tone rules express sequential modifications such as downdrift and coarticulation, glides between discrete tones and in one case replacement (cf sandhi in Thai). Sentence type can be marked locally (sentence tone) by an added L or H, or by replacement, or by overall intonational effects, e.g. downdrift or absence of downdrift, or by combinations of local and overall features. There is no consistent pitch signal for statement or question in these languages. Yet, statement is generally formed by downdrift of the basic tones, whereas for question some sentence-rule has to be added. Exclamations are uniform: a larger range by raising H's. Sentence accent is not mentioned.

Some American Indian Languages, etc. E.V. Pike

The author examines accent (stress) at word and sentence level in nine languages which use two or more contrastive tones. The tonal contrasts occur in both accented and unaccented syllables in eight of the nine languages. One language, Fasú, has tonal contrasts only on accented syllables. Apart from special allotones, such as raised H or lowered L, loudness and length (in consonant or vowel) mark the accented syllable. There are language-dependent rules for the distribution of word and sentence accent. Word accent is connected with the first, last or penultimate syllable of the stem. Sentence accent may fall on the last word accent or on any accented syllable, or it may have a separate manifestation, added to the final syllable. In some of the languages downdrift is reported for statement and downdrift + upglide for question. The prepause syllable has a special tone system for attitudes in two of the languages. Glottal stop, expressing finality, belongs in one of them, along with final-syllable upglide denoting politeness.

Swedish. G. Bruce

The author presents and modifies a model for Swedish intonation (Bruce and Gårding 1978). In this model the word accents (A1, A2) have been analysed as HL's with similar but differently timed Fo-correlates. There are dialectal differences but the HL occurs later in A2 than in A1 in all cases. Sentence accent is manifested as a wide pitch range with dialect-dependent distribution rules. In the main dialect areas a separate H is added after the word accent HL, in other dialects the wide range is obtained together with the accented syllable. Statement intonation is expressed as a progressing downdrift of H's and L's. The author argues that the Fo drop has a stepwise rather than a gradual downdrift. The downsteps cooccur with the accented syllables. Sentence intonation, then, has a systematic influence on Fo-values of H's and L's with higher values for earlier positions in the utterance.

Danish. N. Thorsen

A descriptive model of sentence intonation in Copenhagen Danish is presented. It does not take account of the word accents, stød (A1) and non-stød (A2). Sentence intonation is defined as a line described by the pitch of the accented syllables. An accented syllable and the following unaccented ones form a stress group in which the accented syllable is low in pitch and the unaccented syllables rise above and fall below the sentence-intonation line. This line is steeply falling for declarative sentences and level for unmarked questions. Sentence intonation has a systematic influence on the Fo course in the stress groups, in that the rise from stressed to unstressed syllable is larger in questions than in statements. The author rejects sentence accent but accepts emphatic or contrastive stress with manifestations common to many languages: a raising of pitch on the emphatic syllable at the expense of surrounding accents, i.e. shrinking of pre-emphatic and deletion of post-emphatic pitch movements in connection with accents.

Dutch. 't Hart and R. Collier

Each word has a lexical accent whose location can be predicted by rule. Under the influence of intonation it may be manifested as pitch movement. The authors study the interaction between intonation and accentuation. Two principles are discussed. According to the first, the overall pitch contour is obtained by adding

autonomous accentual pitch movements to autonomous pitch movements associated with sentence type, such as downdrift (declination) for statement and downdrift plus final upglide for question. This is rejected in favour of a second principle according to which the accents only determine the location of the pitch movements. Their nature (rise, fall, etc.) and order are determined by the chosen intonation pattern.

Czech. P. Janota

Janota reports on a series of tests with bisyllabic synthetic stimuli. When both syllables have the same pitch value, the first is judged as accented 85% of the time. A small increase or decrease of  $F_0$  in the second syllable raises the number of accent votes for this syllable from 15 up to about 60%. With a larger change of  $F_0$ , the number of such votes goes down. Responses to similar items, used to evaluate sentence intonation, show that precisely the stimuli with the large  $F_0$  deviations are effective cues to intonation. Stimuli with a moderate or substantial increase of  $F_0$  are judged as continuative statements and questions, respectively, and those with a decrease of  $F_0$  as statements.

Suggested points for discussion

1. Universality of prosodic units

It may be fruitful to accept different degrees of universality, e.g. universal for similar function and similar acoustic correlates for all languages, and near-universal for corresponding conditions in almost all languages. As a third degree I suggest generality, requiring similar function, similar acoustic correlates and many languages.

Sentence intonation is a universal. This may be considered as a postulate. On the other hand, the contributions to this symposium and other communications show that statement intonation expressed as downdrift is merely a generality. The same holds for question intonation expressed as an absence of downdrift (Thorsen, Williamson).

Is sentence accent a universal? Is the lack of sentence accent in the description of some languages (Thorsen, Williamson) due to some possibilities in these languages to express deictic function at the sentence level in a non-prosodic way? Or is their description an artefact due to difference of analysis and tradition?

2. Principles for the analysis of the interaction between sentence prosody and word prosody. In recent work on Swedish intonation, basic forms of prosodic units have been isolated, after a phonological analysis, and rules for their combinations have been formulated (Bruce).

Other methods have also been mentioned. Is it possible to find a common framework applicable to all prosodic systems?

3. Universal features in the interaction between sentence prosody and word prosody.

't Hart and Collier demonstrate that for Dutch, sentence intonation is primary to word prosody. Is this a universal feature if we take primary in the following sense: In production sentence prosody precedes and sets the scale for word prosody. On the other hand, the degree to which word prosody interferes varies from very little as in Czech to something quite drastic as in Thai. Dutch seems to occupy an intermediary position and a 2-accent system like Swedish or Danish is closer to Thai.

Tone rules are a formal convenient way of expressing both co-articulation and the influence of sentence prosody on word prosody (see e.g. Williamson and her references). How universal are the tone rules?

4. Additional questions

Accent or Tone. In her description of various tonal languages Pike mentions one, Fasú, which has tonal contrasts (H, L) only on stressed syllables. What is the difference between a tonal language like Fasú and an accent language like Swedish in which the stressed syllables in some analyses (e.g. Malmberg) also have been represented by H versus L?

Accents and accents. Are there different physiological mechanisms behind different kinds of accent, e.g. the small pitch movements noted in Czech (Janota) and the larger ones typical of Germanic languages (Bruce, Thorsen, 't Hart and Collier). Or are they merely weak and strong manifestations of the same phenomenon?

Note

Two books have just come out which may be relevant for the discussion:

Fromkin, V. (ed.) (1978): Tone. A Linguistic Survey, Academic Press.

Greenberg, J. (ed.) (1978): Universals of Human Language. Volume 2. Phonology, Stanford University Press.

## LEXICAL TONE AND SENTENCE PROSODY IN THAI

Arthur S. Abramson, University of Connecticut, Storrs, Connecticut, U.S.A. and Haskins Laboratories, New Haven, Connecticut, U.S.A.

Background

In a true tone language, one in which, in principle, every syllable in the morpheme stock bears a distinctive tonal phoneme, the tones are characterized primarily by fundamental-frequency levels and contours. Since we also describe intonation mainly in terms of the fundamental frequency ( $F_0$ ) of the voice, there seems to be a paradox involved in examining the relations between sentence prosody and word prosody in a tone language. As in other languages so also in tone languages is there the possibility of expressing attitudes or indicating certain aspects of syntactic structure by means of sentence intonation. The question arises as to whether the effects of this sentence intonation are strong enough to weaken or even destroy the phonetic integrity of lexical tones.

The citation form of a monosyllabic word may be viewed as bearing the ideal manifestation of a tone. Of course, except for the occasional one-word sentence, these ideal forms do not often occur in running speech, yet children in the culture probably learn new words this way, and so do adults in a foreign-language class. Once we have two or more tone-bearing syllables strung together, we expect perturbations through coarticulation. The final physical shaping of a tone is provided by the intonation on the utterance (Pike, 1948, 18-19).

The Tones of Thai

The ideal shapes of the tones of Standard Thai (Siamese) have been described elsewhere (Abramson, 1962; Erickson, 1974). It is useful to divide the five distinctive tones of the language into the "dynamic" class, comprising the falling and rising tones, and the "static" class, comprising the high, mid and low tones. The dynamic tones show rapid movements of  $F_0$ , while the static tones show rather slow movements which sometimes approximate levels. Of the three static tones it is the mid tone that is most likely to appear occasionally as a level. The high tone is more likely to be seen as a rise high in the voice range in contrast with the low rise of the rising tone. The low tone is

likely to appear as a low fall in contrast with the high fall of the falling tone.

Two types of phonetic context perturb the ideal shapes of the tones. Voiceless initial consonants induce a higher start of the  $F_0$  contour, while voiced initial consonants induce a lower start (Gandour, 1974; Erickson, 1974). This kind of perturbation seems to have little effect on the phonetic integrity of the five tones, although it may serve as a supplementary cue to the voicing state of the initial consonant. It has been argued by historical linguists (Li, 1977), with some perceptual support from recent experiments on Thai (Abramson and Erickson, 1978), that through the phonemicization of these perturbations, the tones of Proto-Tai increased from three to the present-day sets of five or more in the modern languages of the family.

The phonetic context that causes greater deviations from the ideal tonal shapes is that of neighboring tones. In a series of tones spoken without pauses, tonal coarticulation occurs. Although physiological studies of Thai tones (Erickson, 1976) have yet to be extended to sequences, we can infer from acoustic evidence (Abramson, 1979) that this kind of coarticulation is manifested through the overlap of the effects of motor commands for the control of the laryngeal tensions and aerodynamic forces used.

Two sequential effects must be discriminated from tonal coarticulation. First, certain unstressed CV syllables with short /a/ which have low or high tones in citation form are normally toneless in running speech. Another view is that the high and low tones on these syllables are neutralized, and the resulting pitch is assigned to the mid tone. This conclusion is handy for transcription, but the physical evidence suggests instability with  $F_0$  values dominated by the contours of the neighboring lexical tones. The other sequential effect to be excluded from consideration is tonal sandhi. The phonology of the language dictates that when certain kinds of morphemes are conjoined to form compound words, the lexical tone of one of the morphemes is replaced by another tone.

Sentence Intonation and Tones

As one listens to spoken Thai, whether it be an animated conversation or a phlegmatic technical explanation, it becomes clear that in addition to emotional states such linguistic features as

sentence accent and signs of major syntactic breakpoints can be expressed prosodically. The distinction between a statement and a question can also be expressed. In my present approach to the topic, I must lean mainly on my own extensive auditory but limited instrumental observations, as very few useful insights are found in the literature. It would be helpful if native Thai linguists or phoneticians gave more attention to the matter.

As a data base for such observations, as I am ready to make, I have used two kinds of speech material. One is a conversation between two Thai adults of about one minute in length, recorded by J. Marvin Brown for a textbook published by the American University Alumni Association Language Center in Bangkok, Thailand. The other is a monologue recorded by me of the dean of a faculty at a university in Bangkok; speaking for a bit more than a minute, he talks about a new academic program.

Computer-implemented analysis yielded displays of root-mean-square amplitude, wave forms, and  $F_0$  contours. Cepstral analysis was used to extract the fundamental frequency. A sample set is shown in Fig. 1 for the female speaker in the dialogue. Here, by the way, can be seen an example of tonal coarticulation. The phrase /nǎ bǎn/ 'in front of the house' bears two falling tones. The  $F_0$  of the first one does not fall as far as the second; this presumably facilitates the resetting of the larynx for the sharp rise and fall of the second falling tone.

To handle the non-emotive aspects of sentence prosody in Thai, my examination of the present corpus of utterances, reinforced by the arguments of Rudaravanija (1965), leads me to posit three terminal junctures: rising pitch, sustained pitch, and falling pitch. These junctures function at clause ends and sentence ends. They may also function wherever the speaker pauses. The presence of a juncture affects the phonetic shape of the lexical tone on the last one or two syllables. The rising and falling junctures are likely to appear at the end of a breath group. In earlier work (Abramson, 1962) I also posited two pitch registers, high and normal, as units for Thai intonation. I now doubt the relevance of such registers for the non-emotive aspects of sentence prosody in the language. Indeed, to capture emotive prosodic variation, a somewhat more elaborate scheme might be needed. Although, as shown by Noss (1972) and Thongkum (1976),

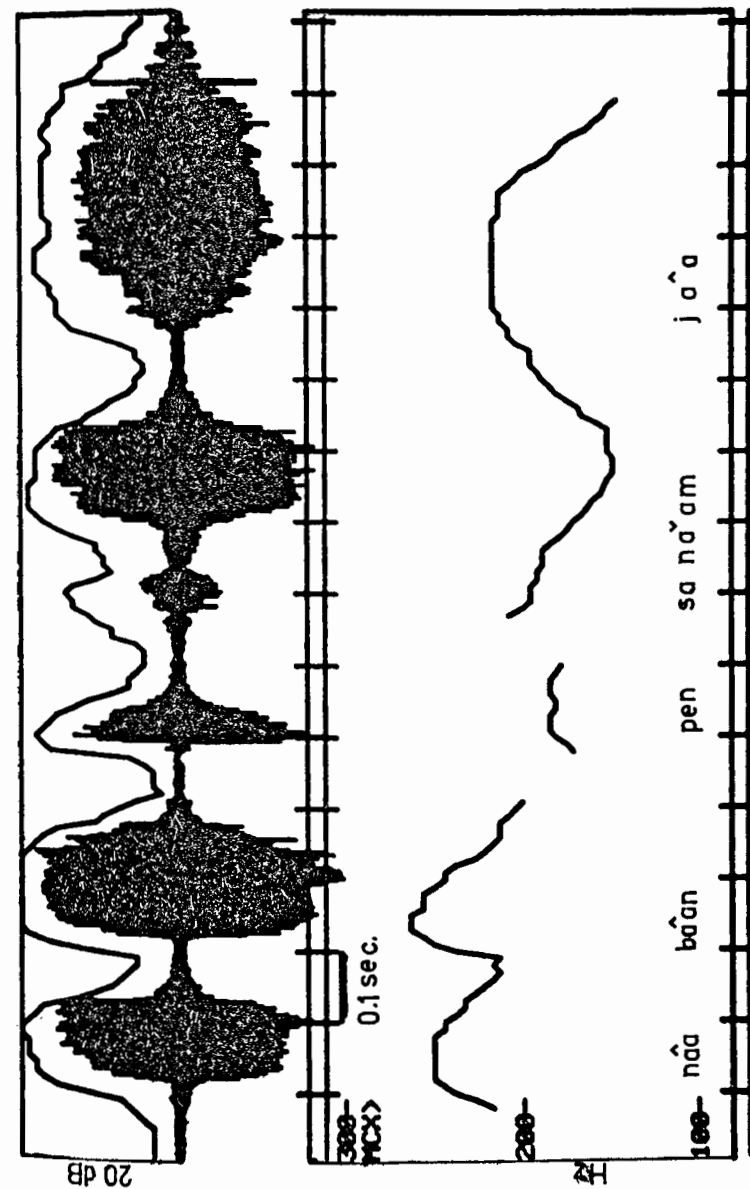


Fig. 1. From top to bottom: R.M.S. amplitude, wave form and fundamental frequency of a Thai woman's production of the sentence /nǎ bǎn pen sa nǎm jǎ/ 'There is a lawn in front of the house.'



rhythmic factors play a role in Thai sentence prosody, they are excluded here because of the scope set by the organizers of the Congress.

Henderson (1949) has argued that aside from the general melodic line of Thai intonation, the "sentence tone" as a whole is mainly determined by the speaker's choice of particles, most of them final particles. She describes seven such sentence tones. Without entering into the question of how many sentence tones there might be, I can at least say that these particles, which indicate, e.g., the sex of the speaker and something about the social relation between the speaker and the hearer, are prime carriers of the terminal junctures. Each particle as a lexical item has a tone of its own in citation form; this tone is usually predictable from the spelling. I doubt, however, that in running speech these "lexical" tones have any standing. The actual pitch imposed on a particle or, sometimes, a sequence of two particles, seems to be determined by the intonation of the whole sentence culminating in a terminal juncture. The resulting "tones" on these particles can sometimes be aligned with the lexical tones of Thai phonology but more often they are deviant; some linguists, apparently in the grip of the view that every Thai syllable must bear a phonemic tone, feel constrained to write each particle with one of the five tones.

In both colloquial and formal discourse, many a sentence contains no particles, so the terminal junctures appear on the final word of the clause or sentence. Fig. 1 shows such an effect. The falling tone on /jɑ̀/ 'grass' at the end of the sentence is considerably lower both at its high point and low point than the two falling tones at the beginning. Even the rising tone just before it on /sanɑ̀m/ 'field' does not rise to a point much higher than the immediately preceding mid tone on /pen/ 'be'. With such a short utterance it is hard to decide whether we have a final falling juncture on the compound word for 'lawn' or a falling intonation contour on the whole sentence.

Sentence accent is manifested by one or more of the following factors: (1) lengthening of the syllable, (2) a tonal contour that approaches the form of the ideal tone, and (3) an increase in amplitude. In the sentence in Fig. 1 the final syllable appears to bear the sentence accent, using factors (1) and

(2). In the phrase /nà bàn/ at the beginning of the sentence, the second syllable is stressed, using factors (1) and (3); the amplitude trace is flattened at the top of the available 20-dB range, indicating saturation.

The points made so far have been descriptions of gross  $F_0$  contours. A problem in intonation analysis is how to present quantitative data that go beyond overall "tunes." The prosodic constructs of the linguist often elude the measuring devices of the phonetician. With the simple-minded analysis for non-emotive prosody into three terminal junctures as a framework, I have made an initial tabulation of frequency movements for such clear examples of terminal juncture as I could find in the corpus. To provide for reasonable comparability of speakers, I treated frequency shifts at terminal junctures as percentages of the voice range. The maximum and minimum  $F_0$  values for each of the three speakers are given in Table 1. Although the speech in both samples was

Table 1

	Voice Range in Hz		
	Dialogue		Monologue
Speakers:	A*	B**	U.W.*
Spread:	130-290	90-235	85-160
Range:	160	145	75
	*Woman	**Man	

calm, the narrower range for the monologue may not be due so much to the habits of that speaker as to the rather dispassionate and thoughtful nature of his discussion compared to the more animated dialogue.

The juncture of sustained pitch is generally found at syntactic breaks where the overall pitch of the voice neither rises nor falls before a brief pause; with or without a pause, the final syllable is prolonged. I have used this sustained pitch as a neutral reference from which to track the movements of the other two junctures. Examining both samples by ear and by eye, I accepted as valid tokens of the three junctures only those instances that were quite unambiguous. This cautious procedure yielded the small number of data in Table 2. The juncture of rising pitch signals surprise, doubt or a question. (Questions can also be marked by means of particles and other morphemes without terminal rising pitch.) The terminal fall appears at the ends of sentences



Table 2  
Average Shift Through Voice Range for Terminal Pitch Junctures

Rising		Sustained		Falling	
N	%	N	%	N	%
6	30	14	0*	27	25

\*Neutral reference point.

and some major clauses. The "shift" for the sustained pitch is set at 0% as a neutral reference level, while the other two junctures are entered as departures from that neutral level. The data are averaged across the three speakers. None of the tokens of these junctures happened to occur with the low lexical tone.

Even away from the junctures intonation has great effects on the realizations of the tonal phonemes. If the ideal forms of the tones have any psychological validity, then the forms in the sample of running speech have undergone severe distortion. A full account is beyond my reach here. At the same time, as I look at the contours and listen to the speech, I find preservation of the full system of five tones in running speech. That is, the usual linguistic scheme is not an artifact of the formal analysis of the linguist concentrating on citation forms only. Excluded from this generalization, however, must be all particles occurring at major syntactic breaks; they generally have their pitch determined by the sentence intonation without the involvement of lexical tones. Other frequently used function words, such as modals and pronouns, often undergo tonal replacement.

#### Conclusion

The phonemic tones and sentence prosodies of Thai interact in a rather complicated fashion. Three terminal pitch junctures, often occurring on particles, carry much of the intonation. Although the lexical tones are much influenced in their  $F_0$  movements by sentence intonation, the contrasts between them are preserved except for certain small sets of morphemes. Sentence prosody allows for sentence accent. As in non-tonal languages, it is possible in Thai to use pitch junctures for the difference between statements and at least some kinds of questions.

#### References

Abramson, A. S. (1962): The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments, Bloomington, Indiana: Indiana U. Res. Center in Anthropology, Folklore and Linguistics.

- Abramson, A. S. (1979): "The coarticulation of tones: an acoustic study of Thai", in Studies in Tai and Mon-Khmer Phonetics in Honour of Eugenie J. A. Henderson, V. Panupong, P. Kullavanijaya and T. Luangthongkam (eds.), Bangkok: Indigenous Langs. of Thailand Res. Proj.
- Abramson, A. S. and D. M. Erickson (1978): "Diachronic tone splits and voicing shifts in Thai: some perceptual data", Haskins Labs. Status Report on Speech Research SR-53, Vol. 2, 85-96.
- Erickson, D. (1974): "Fundamental frequency contours of the tones of Standard Thai", Pasaa 4, 1-25.
- Erickson, D. M. (1976): A Physiological Analysis of the Tones of Thai, Ph.D. diss., University of Connecticut.
- Gandour, J. (1974): "Consonant types and tone in Siamese", JPh 2, 337-350.
- Henderson, E. J. A. (1949): "Prosodies in Siamese: a study in synthesis", Asia Major N.S. 1, 189-215.
- Li, F.-K. (1977): A Handbook of Comparative Tai, Honolulu: U. Press of Hawaii.
- Noss, R. B. (1972): "Rhythm in Thai", in Tai Phonetics and Phonology, J. G. Harris and R. B. Noss (eds.), 33-42, Bangkok: Central Inst. of English Language.
- Pike, K. L. (1948): Tone Languages, Ann Arbor, Mich.: U. of Michigan Press.
- Rudaravanija, Panninee (1965): An Analysis of the Elements in Thai that Correspond to the Basic Intonation Patterns of English, Ed. D. diss., Teachers College, Columbia U.
- Thongkum, T. L. (1976): "Rhythm in Thai from Another View Point", Pasaa 6, 144-158.

## WORD PROSODY AND SENTENCE PROSODY IN SWEDISH

Gösta Bruce, Phonetics Department, Lund University, Sweden

Introduction

The present paper summarizes the research on Swedish prosody reported in Bruce (1977) and Bruce and Gårding (1978), and also presents some new ideas concerning intonation in Swedish. The main topic is the relation between word accent, sentence accent and sentence intonation as signalled by  $F_0$ . Our results suggest that observed  $F_0$ -contours typical of statements in four prosodically distinct dialect types (see e.g. Gårding 1975) represent the combined result of one common sentence intonation command, similar word accent commands with different timings, and different sentence accent commands.

Sentence accent

In a prosodic typology for Swedish dialects combining word and sentence prosody (Bruce and Gårding 1978) we have shown how four prosodic dialect types (south, central, east and west) can be characterized mainly by differences in sentence accent. This is illustrated in Figure 1, which shows  $F_0$ -contours of the sentence Man vill lämna våra långa nummer (They want to leave some Långa-numbers), where the placement of sentence accent has been varied. The two accented syllables (accent II) in the sentence - belonging to the verb and the nominal compound respectively - and the secondary-stress syllable in the compound are all surrounded by unstressed syllables. We regard this form of the sentence as optimal for revealing prosodic dialect differences, since the tonal commands can be developed freely with no obvious interference from adjacent commands.

In the south and central dialects, the sentence accent command appears to be superimposed on the word accent command (see Figure 1). Sentence accent is signalled by a wider  $F_0$ -range for the word accent in focus than for the non-focal word accent. For south this wider range is obtained in final position by raising the actual word accent peak, and in non-final position also by lowering the subsequent valley. For central, the wider  $F_0$ -range is achieved mainly by raising the word accent peak in focus both in final and non-final position.

For east and west the sentence accent command comes after

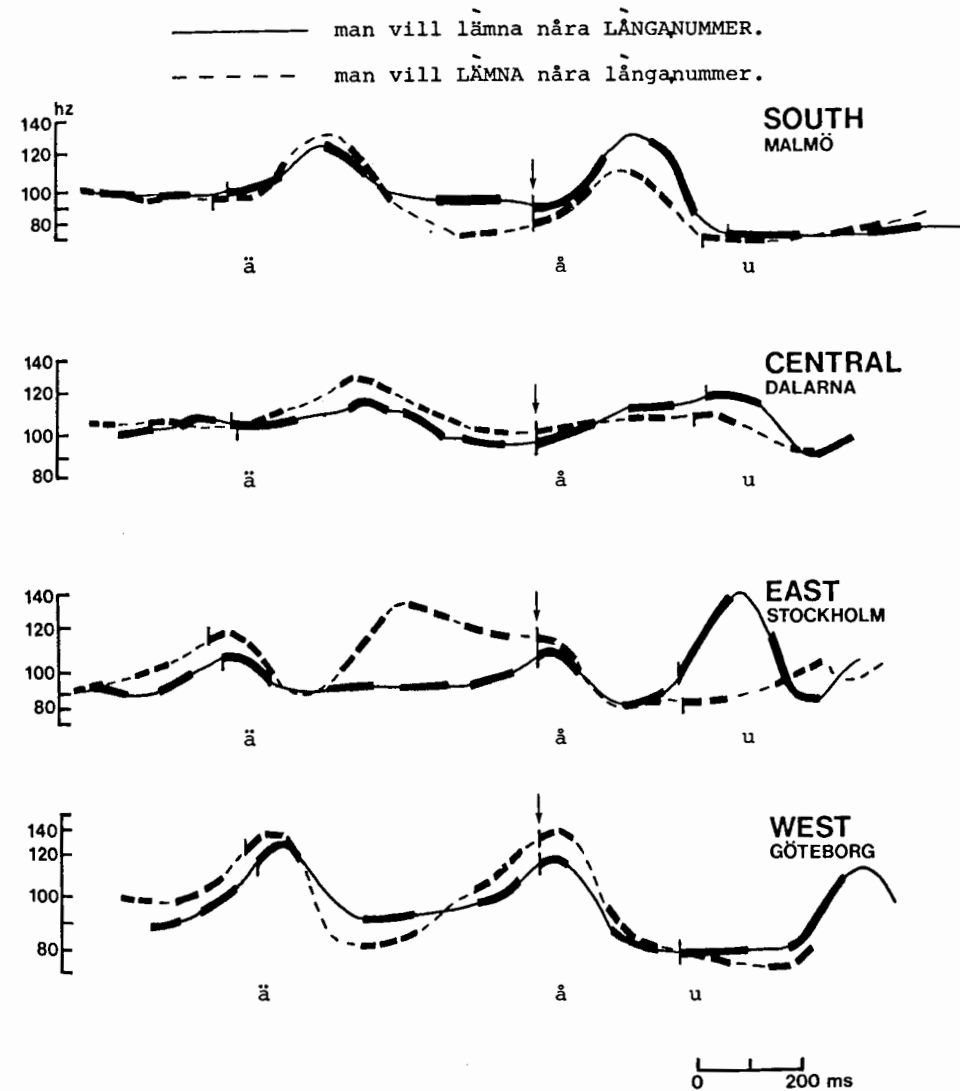


Figure 1. The effect of the placement of sentence accent.  $F_0$ -contours of a sentence with two accented words (accent II). The line-up point (arrow) is at the CV-boundary of the second stressed syllable. Vertical bars indicate the CV-boundaries of the stressed syllables. Vowel segments are drawn in thick lines and consonant segments in thin lines. Focus is indicated by capital letters and non-focus position by small letters.

the word accent command. For east, this means the addition of a separate sentence accent peak immediately after the word accent in focus. For a compound, however, the second peak is postponed until the secondary-stress syllable. The Fo-peak may become a plateau, since in final position Fo will stay on a high level until the utterance-final syllable and in non-final position until the following accented syllable.

For west, the addition of a sentence accent peak in final position occurs in the utterance-final syllable independently of the prosodic structure of the word in focus. In non-final position sentence accent is signalled by both lowering the valley after the word accent in focus and raising the peak of the post-focal word accent, i.e. by a wider Fo-range after the word accent in focus.

This means, in summary, that sentence accent is characterized by a wide Fo-range in all dialect types. In south and central this wide range cooccurs with the word accent in focus, while in east and west it occurs with a time lag.

#### Word accent

When the contribution of sentence accent to the Fo-contour has been isolated, the word accents appear more clearly. Figure 2 shows Fo-contours of the two word accents (accent I and accent II)

in a non-focal position. It appears that for each dialect type the word accent distinction is signalled mainly by the different timing of the word accent peak relative to the stressed syllable (cf. Haugen 1949). The relative timing difference with accent I always preceding accent II is common to all dialects, but the absolute timing of the word accent peaks varies with dialect. The order of timing between dialects from the earliest to the latest absolute timing of the word accent peaks is east, west, south and central (see Figure 2). For east, at the one extreme, the accent I-peak appears as early as in the pre-stress syllable, while for central, at the other extreme, it occurs in the final part of the stressed syllable. The accent II-peak occurs for east in the initial part of the stressed syllable and for central in the post-stress syllable. The timing of the word accent peaks for west and south falls in between these two extremes.

#### Sentence intonation

While the dialect-specific features of Swedish intonation

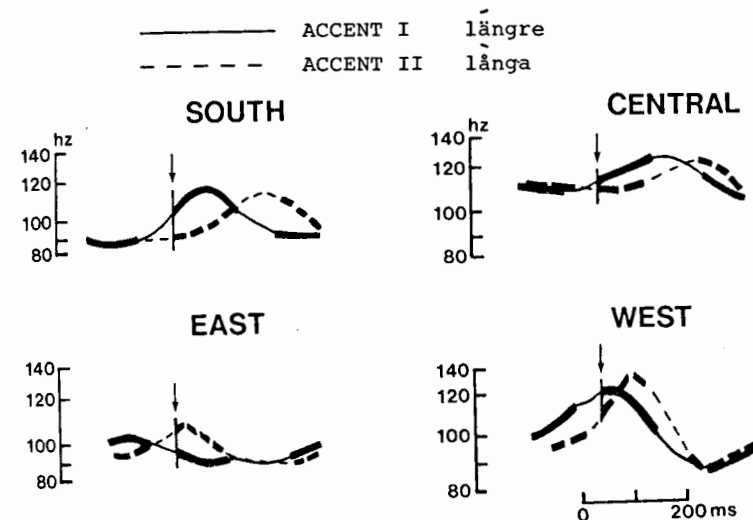


Figure 2. The word accent distinction in non-focal position. Fo-contours of accent I and accent II before focus. The line-up point is at the CV-boundary of the stressed syllable.

are mainly found in the Fo-correlates of word accent and sentence accent, certain aspects of sentence intonation seem to be independent of dialect. Here only statement intonation will be treated. But it appears that also the main features of question intonation are more or less dialect-independent (see Gårding forthcoming).

A characteristic feature of sentence intonation in many languages of the world is the downdrift of Fo over the course of an utterance, also referred to as the declination effect (see e.g. Cohen and 't Hart 1967). This means, in general, that Fo is higher in the beginning than at the end of an utterance with each Fo-peak and Fo valley being lower than the preceding one. In Swedish the topline connecting successive peaks of an utterance appears to decrease at a faster rate than the baseline connecting successive valleys, which means that the Fo-range is also gradually decreasing. This Fo-downdrift is found in English and Danish, too (see Breckenridge 1978, Thorsen 1978).

A model was proposed in Bruce and Gårding (1978) to account for the main features of sentence intonation in Swedish. In my experience it is not typical of the Fo-downdrift in Swedish to be

evenly distributed over an utterance. The total Fo-drop in an utterance for a given speaker appears to be the same, however, regardless of the length of the utterance. The actual course of the Fo-downdrift in an utterance seems to be dependent on several factors, such as the location of sentence accent and of the word accents. Figure 3 illustrates this point. It shows Fo-contours of the sentence *Man vill lämna våra långa nunnor* (They want to leave some tall nuns) containing three accented syllables (accent II) with varying placement of sentence-accent.

The Fo-drop appears to have a stepwise and not gradually decreasing course. The downstep takes place in connection with the accented syllable. In unaccented syllables before and after a word accent there is no systematic downward slope.

It will be assumed that the basic sentence intonation command (statement intonation) has a stepwise decreasing course with a successive narrowing of the range. Sentence accent normally interferes with sentence intonation, introducing a break into the basic pattern (see Figures 1 and 3). This may affect the course of the topline as well as that of the baseline depending on the dialect. Before focus the Fo-drop and the narrowing of the Fo-range of the word accents appear to be relatively gentle. After focus, however, it usually decreases more rapidly, with a considerable narrowing of the Fo-range.

As a consequence of the Fo-downdrift there is a position-dependent variation of word accent and sentence accent. A word-accent in the beginning of an utterance has higher Fo-values for peak, valley, and range than at the end (everything else being equal). Also the corresponding sentence accent values tend to decrease with position in the sentence.

#### Conclusion

Sentence intonation (statement intonation) in Swedish can be represented independently of dialect by a stepwise decrease of Fo taking place in connection with the word accents and affecting peak, valley, and range values. This downdrift pattern will often be locally disturbed by sentence accent introducing a break into the basic Fo-course in a dialect-specific way. This suggests that the downdrift is linguistically controlled rather than a consequence of some peripheral production constraint. Instead it can be assumed to be built into the intonation system.

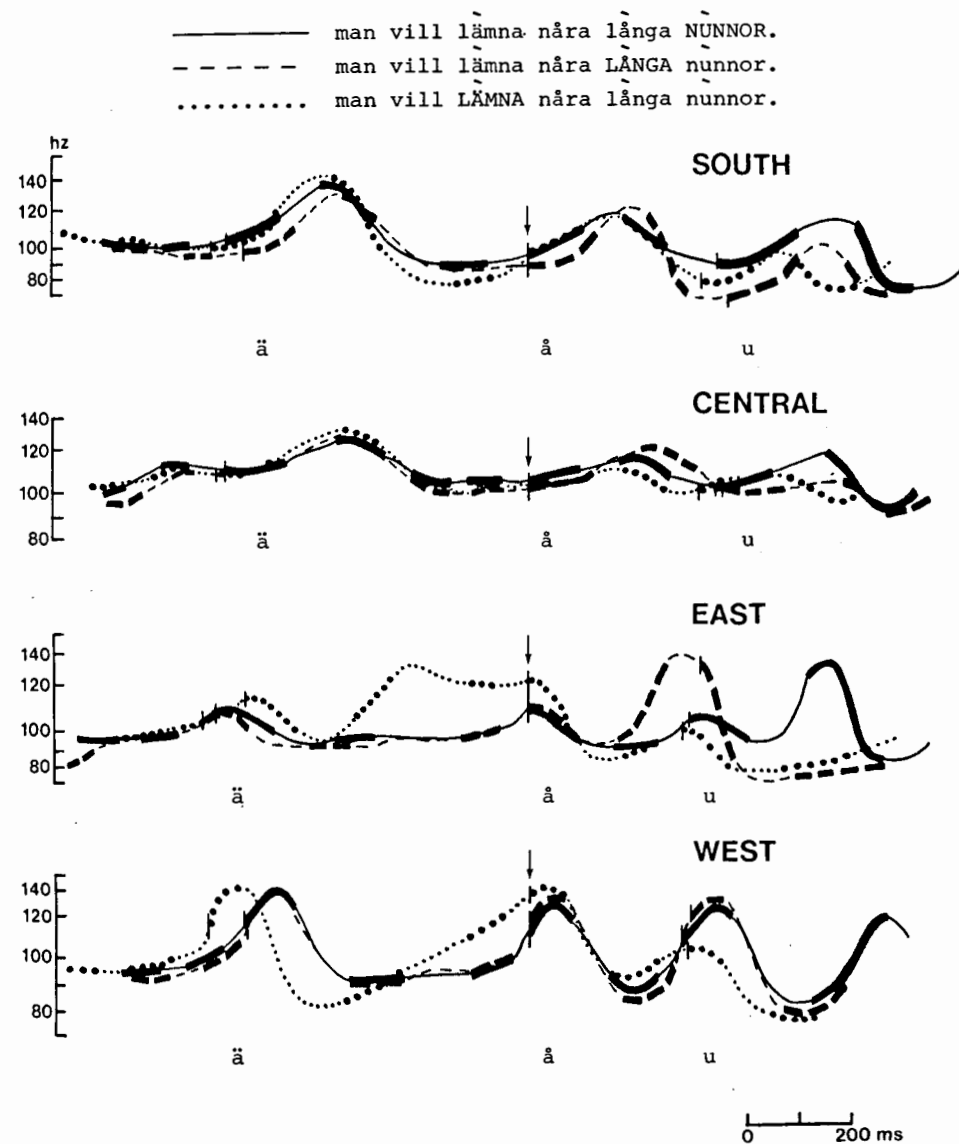


Figure 3. Downdrift in Swedish - the combined effect of statement intonation and the location of sentence accent and word accent. Fo-contours of a sentence with three accented words (accent II). The line-up point is at the CV-boundary of the second stressed syllable.

Acknowledgements

This work was carried out within the project "Swedish prosody" in cooperation with Eva Gårding. It was sponsored by the Swedish Humanistic and Social Sciences Research Council.

References

- Breckenridge, J. (1978): Declination as a phonological process, Department of Linguistics and Philosophy, MIT, Cambridge, Massachusetts, Mimeographed.
- Bruce, G. (1977): Swedish word accents in sentence perspective, Lund: Gleerup.
- Bruce, G. and E. Gårding (1978): "A prosodic typology for Swedish dialects", to appear in Nordic prosody, E. Gårding (ed.), Lund: Gleerup.
- Cohen, A. and J. 't Hart (1967): "On the anatomy of intonation", Lingua 19, 177-192.
- Gårding, E. (1975): "Towards a prosodic typology for Swedish dialects", The Nordic languages and modern linguistics 2, 466-474, K.-H. Dahlstedt (ed.), Stockholm: Almqvist & Wiksell.
- Gårding, E. (1977): The Scandinavian word accents, Lund: Gleerup.
- Gårding, E. (forthcoming): Sentence intonation in an accent language.
- Haugen, E. (1949): "Phoneme or prosodeme?", Language 25, 278-282.
- Thorsen, N. (1978): "Aspects of Danish intonation", to appear in Nordic prosody, E. Gårding (ed.), Lund: Gleerup.

## ON THE INTERACTION OF ACCENTUATION AND INTONATION IN DUTCH

J. 't Hart and R. Collier, Institute for Perception Research,  
P. Box 513, Den Dolech 2, Eindhoven 5612 AZ, The Netherlands

Introduction

The general aim of this symposium is to track down universal features concerning the relation between word prosody and sentence prosody, with the exclusion of durational phenomena. We will present our viewpoints about the relation at issue in as far as we have been confronted with it in our experiences with Dutch intonation and accentuation.

Dutch intonation lacks the occurrence of "tonemes". Therefore, on the level of the word, we can limit our discussion to phenomena of "lexical accentuation"; on the level of the sentence we have to discuss "sentence accents" and "intonation". On the latter level particular problems may arise as regards the interaction of accentuation and intonation. In fact, sentence accents manifest themselves as "pitch accents" on words or syllables, whereas intonation patterns are realized as "pitch contours" extending over entire utterances. In other words, two aspects of sentence prosody are interwoven in the same phonetic variable, viz. the variation of  $F_0$  (or pitch) as a function of time.

In the literature, the problem of the interaction between accentuation and intonation is coped with in a number of ways, most of which share the assumption that the overall pitch contour can be considered simply as the sum, or the linear addition, of the variations of  $F_0$  associated with accentuation and those associated with intonation. We will confront this assumption with a number of phenomena as observed in Dutch.

Word prosody

In Dutch, each word is said to possess a lexical accent. In polysyllabic words the location of this accent is not fixed but, in principle, it can be predicted by rule.

Lexical accents may be considered as abstract features on the level of word phonology. A subsequent and separate problem is to specify when and how these lexical accents manifest themselves phonetically.

If we take, for example, the word Amerika (America), we find a lexical accent on the second syllable. When this word is spoken in isolation, as a one-word utterance, a listener will hear the

second syllable as prominent. Acoustic measurement of the utterance will reveal substantial changes of  $F_0$  (hereafter referred to as "pitch movements") on the second syllable: a rise, a fall, or a combination of the two. These pitch movements may be considered a phonetic correlate of the lexical accent, since it can be shown experimentally that their deletion or displacement causes prominence judgments to change accordingly. That pitch movements are efficient cues for prominence is not surprising if one realizes that, psycho-acoustically speaking, only a few percent change of  $F_0$  is sufficient to be supraliminal, whereas in actual speech  $F_0$  changes are observed that are eight to ten times as large as these threshold values.

At this point we may conclude that lexical accents in one-word utterances are realized (among other things) by means of pitch movements: rises, falls, or combinations of both, apparently without a particular preference for any of these various possibilities.

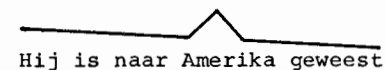
Since a one-word utterance is an utterance all the same, we may as well extend the discussion to longer utterances with one or more accents.

#### Sentence prosody

In the introduction we have already mentioned that on the level of the sentence, pitch changes correlate with both accentuation and intonation. On the linguistic level these categories are easily kept apart, but on the phonetic level the distinction may become blurred to the extent that the observable pitch changes can be associated with either of the two categories (or with both). One would therefore like to sort out how accentuation and intonation interact in shaping up the ultimately observable course of the pitch in concrete utterances.

Let us illustrate this problem with the following example. The Dutch sentence Hij is naar Amerika geweest (He has been to America) may be pronounced with one accent, viz. on the syllable -me- of Amerika. Again,  $F_0$  measurements will show pitch movements on that syllable, e.g. a rise-fall combination. Of course, a sufficiently refined  $F_0$  measurement will also reveal changes on other syllables than the accented one, but experiments with artificial, stylized pitch contours show that such changes are not

relevant to the perception of either accentuation or intonation. In the example below, the stylized rise-fall may be preceded and followed by a gradual downward running movement of pitch (the so-called "declination"). This is sufficient to make the contour prosodically well-formed.



The essential shape of the pitch contour of this example corroborates our prior suggestion that, phonetically, no distinction can (nor has to be) made between the realization of a lexical accent in a one-word utterance and the realization of a single accent in a longer utterance: both may manifest themselves in the form of the same pitch movement(s). In both cases the pitch movements are the phonetic correlate of the accent inasmuch as their deletion or displacement has consequences for the perception of prominence.

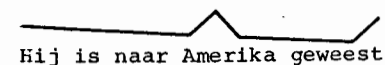
#### Primacy of accentuation

Let us now consider one possible explicit formulation of the principle that the overall pitch contour is obtained by a linear addition of the accentual and intonational requirements. The principle would then be phrased as follows:

P1 Those pitch movements that co-occur with prominent syllables are entirely and exclusively related to accentuation, the remaining pitch movements of the contour are associated with intonation.

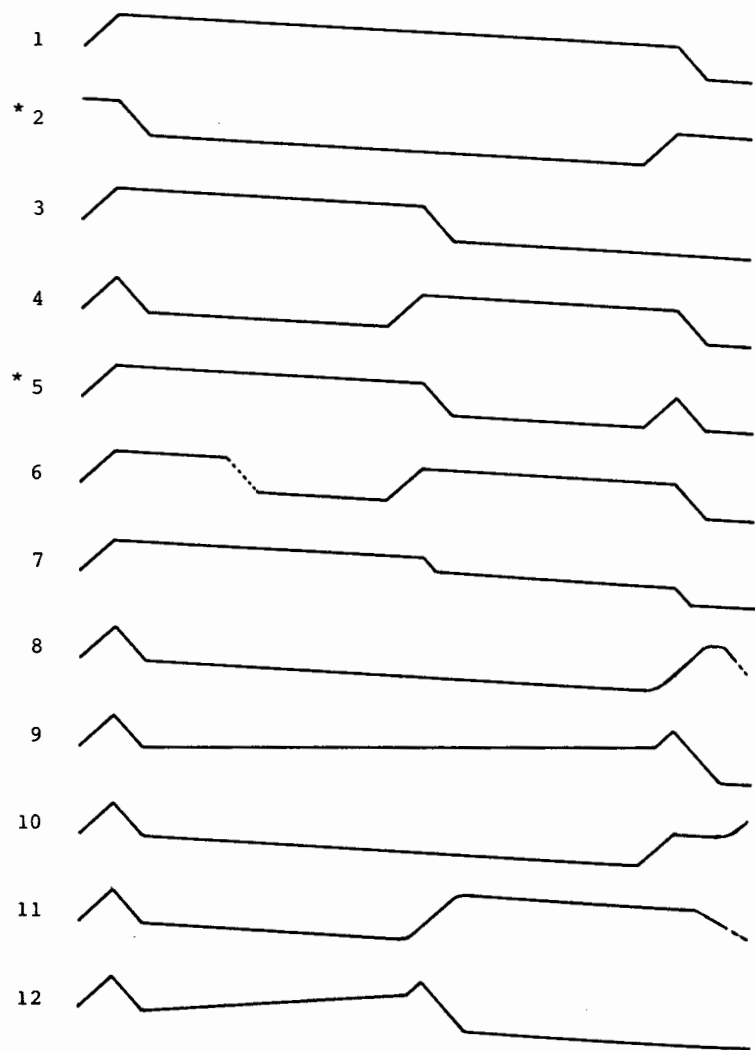
How would P1 account for the pitch contour of Hij is naar Amerika geweest? P1 might relate either the rise or the fall to accentuation and the other pitch movement (plus the declination) to intonation; or P1 would assign the rise-fall combination to accentuation, in which case there would be no particular intonation (except for the declination).

The sentence Hij is naar Amerika geweest may also be pronounced as follows:



In this instance, P1 would assign the final rise to intonation,

Variant:



Grootmoeder gaat met de kinderen naar het zwembad

Fig. 1. Stylised representation of twelve different pitch contours for a given specimen sentence. Variants 2 and 5 are not acceptable.

since it does not lead to an additional accent.

Let us now turn to another, more complicated Dutch specimen sentence, with variants of accentuation and intonation. These variants are listed in Figure 1. They all refer to possible ways of intoning the sentence Grootmoeder gaat met de kinderen naar het zwembad (Grandmother goes with the children to the swimming-pool).

In variant 1 the syllables groot- and zwem- are to be accented. This can be brought about by having the pitch go up on the syllable groot- and down on the syllable zwem-.

Since, apparently, it is possible to produce accent by a mere rise or fall, we might examine the case of variant 2, with a fall on the syllable groot- and a rise on the syllable zwem-. Such a contour can easily be constructed and made audible by means of a speech synthesizer or an Intonator. It appears that the contour of variant 2 sounds unacceptable to Dutch ears. This means that the choice of the kind of pitch movement that has to take care of accentuation is not free.

P1, then, is unsatisfactory since it does not account for the choice of the accent lending pitch movements. There is nothing in the nature of the accents themselves that would predict this choice. So, our suggestion is that the choice of the kind of accent-lending pitch movements is subordinate to the kind of intonation pattern that is to be realized.

#### An alternative: primacy to intonation

This suggestion of subordination is not reflected in the formulation of P1. On the contrary, P1 assumes a primacy in favour of the pitch movements needed for accentuation. A logical alternative would take the primacy of intonation as a starting-point. Basing ourselves on this primacy, we will first formulate an alternative principle, P2, and then try to present empirical support for it.

- P2
- (a) The nature and the order of all the pitch movements in an utterance are determined by the intonation pattern.
  - (b) Among the pitch movements of any intonation pattern there is at least one which possesses such phonetic properties as are necessary for bringing about a pitch accent.
  - (c) The location of the accent-lending pitch movement(s) is



determined by the position of the words that carry sentence stress, and more specifically, by the position of the lexically accented syllable in each of these words.

This formulation does not allow any movement to be entirely and exclusively related to accentuation. Therefore, the addition principle is abandoned.

Returning to variant 1, we would, on the basis of P2, interpret its pitch contour as follows: the intonation pattern which is realized consists essentially of an accent-lending rise and an accent-lending fall, in that order; their locations are in accordance with the accentual demands.

We will now check whether such an account would also be applicable to other accentual and (or) intonational variants.

P2 would predict that if the accentuation requirements change, the only change(s) will be in the location of the accent-lending pitch movements. Suppose, for instance, that the same intonation pattern as in variant 1 is used, but that instead of the syllable zwem-, the syllable kin- has to be accented. P2 will predict a rise and a fall in that order, the rise on groot- and the fall on kin-. This gives rise to variant 3, which is indeed a possible, and well-formed contour.

A special case is provided in sentences with only one accent. If still the same intonation pattern is to be used, then the rise and the fall must necessarily coincide on the one accented syllable. This accounts for the example Hij is naar Amerika geweest.

Yet another illustration of the same intonation pattern with different accentuation is shown in variant 4, where three accents are at stake. In such a case, the introduction of one additional accent-lending pitch movement would in principle be sufficient. However, since the essential property of the intonation pattern being used is "a rise followed by a fall", there is no other possibility than a repetition of these two movements. These may be combined on one of the accented syllables, but not just on any of them. Indeed, an intonational requirement in Dutch is that the separated rise and fall should only occur on the penultimate and the last accented syllable, respectively. This requirement is violated in variant 5, which is therefore unacceptable.

The rise and fall that were introduced in variant 4 to account

for the additional accent need not coincide on the same syllable. The accentuation requirement can also be met by means of a mere rise on the syllable groot-. The fall may then occur somewhat later, as in variant 6. In such a case it coincides with a word boundary (and more specifically with a major syntactic boundary). The fall at the word boundary cannot give rise to an additional pitch accent, due to its particular location, but it may serve another purpose, viz. the marking of a syntactic boundary.

All this goes to say that the intonational requirements are met in such a flexible way that the accentual (and sometimes also syntactic) demands are satisfied at the same time. P2 accounts for this flexibility.

We have shown before that P1 cannot account for the unacceptability of variant 2. P1 would be capable of accounting for the structure of variants 3, 4, 5 and 6. But it cannot explain why variant 5 is unacceptable, nor why variants 4 and 6 are melodically dissimilar, while remaining accentually identical.

If P1 is unsatisfactory to a limited extent in the case of a number of accentual variations in contours based on one intonation pattern, its complete failure becomes apparent whenever variations of the intonation pattern are at stake. This can be illustrated by means of variants 7 to 10, in which four different intonation patterns are used, again with accents on the syllables groot- en zwem- (and in variant 7 also on kin-). P2 accounts for the variants 7 to 10 by stating that when different intonation patterns are realized, some of their essential components fall into such places as is necessary to accommodate the pitch accents. P1, however, could never explain how the accent on e.g. the syllable zwem- is phonetically manifested in so many mutually exclusive ways. Again, if the location of one of the accents changes, the accent-lending pitch movement that figures among the indispensable components of the intonation pattern is shifted to a different position, as appears from the comparison of variant 11 to 8, and of 12 to 9.

#### Conclusion

The examples given show that pitch accents can have quite a number of different forms of appearance. If, according to P1, one would assume that the pitch movements associated with accentuation

are entirely autonomous and do not, in any respect, constitute a part of the essential components of the intonation pattern, this leads to the following difficulty: how will one predict for every single accent which type(s) of pitch movement should be used to realize it? The ability to make such predictions is a real necessity since the various kinds of pitch movement cannot follow one another in an arbitrary order in an utterance.

Therefore, the most satisfying way to account for the combinatory restrictions among pitch movements associated with accents is to assume that they are determined by the chosen intonation pattern. This is expressed by P2, in which, contrary to P1, accentuation is subordinate to intonation.

In our opinion, P2 also accounts for the interaction between pitch accents and intonation patterns in other "accent languages" without tonemes, such as English, German, and others.

## SOME OBSERVATIONS ON THE PERCEPTION OF STRESS IN CZECH

Přemysl Janota, Institute of Phonetics, Charles University,  
Prague, Czechoslovakia

Considerable attention has been paid to the problems of the perception of stress during the last three decades. Generally speaking, the results of the various experiments have largely coincided in showing that there is no one-to-one correlation between judgments of stress and any single physical feature of the speech signal.

In an earlier experiment we tried to determine the influence of three physical dimensions - intensity, fundamental frequency, and duration - on the perception of stress in Czech; the sound material consisted of synthetic disyllabic items. In listening tests of this kind, the listeners' judgments are based, generally speaking, on two complexes of phenomena, which may be labelled as, firstly, acoustic properties of the speech signal and, secondly, contextual cues. The relation between the two complex factors can vary within considerable limits in natural utterances; in experimental conditions it is possible, however, to suppress, to a certain extent, the part played by one of the factors. We also treated the selection of a suitable test word as important. In the experiments we used throughout combinations of the syllable 'se': the word 'sese', meaning 'session', does exist in Czech, though it is quite rare - this means, it is less likely to evoke the association of a word with first syllable stressed, and even less likely in the light of the fact that an actual sequence of syllables se-se is very common in Czech.

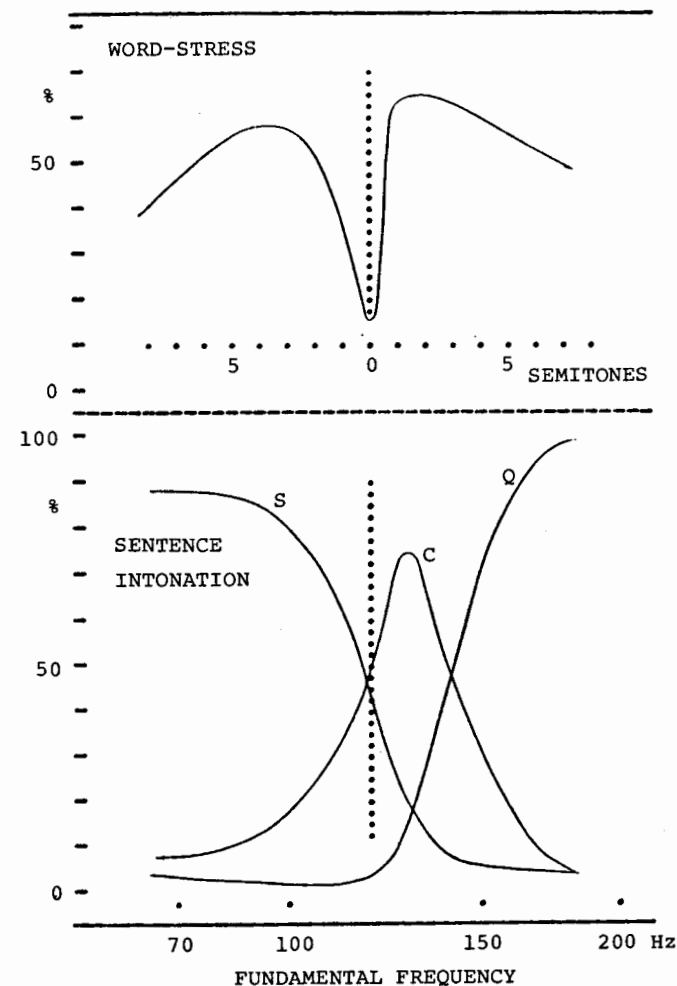
In the instructions to the test, the listeners were invited to mark stressed syllables; in English terminology the term 'prominence' would perhaps be more fitting, but the corresponding term does not exist in Czech phonetic literature. In the last edition of the standard handbook of Czech orthoepy (Hála 1967, 65) it is stated: 'By stress we usually understand the phonetic emphasis of one syllable with respect to others; within a single word this emphasis is called word-stress.' The task of the listeners was then to mark this 'phonetic emphasis', prominence. The test stimuli were produced by means of a simple synthesizer of our own construction, recorded on tape in random order and presented to 170 Czech

listeners. The results of the test can be summarized as follows: increasing intensity of the stimulus leads to an increase in the number of 'stressed syllable' judgments; increased duration has a similar effect. With changes of the fundamental frequency the relationship was found to be somewhat different. Relatively small changes - a semitone up or down - led to a conspicuous increase in the number of judgments 'stressed syllable', while a further raising or lowering of the fundamental frequency did not lead to any further increase in the number of 'stressed' judgments, but on the contrary led to a decrease.

In the evaluation of the results of this experiment we were aware that they were valid for the experimental conditions of the test and for the synthetic material used. However, we have left as an open question the unexpected influence of frequency changes on stress evaluations, i.e. the stronger effect of small changes of the fundamental frequency and the similar effect of both the increase and decrease of the fundamental frequency on the number of judgments 'stressed'. One hypothesis here was that more marked changes of tonal pitch are evaluated rather as sentence melody. Besides, a possible influence of the synthesizer had to be taken into consideration.

In the following tests we concentrated specially on these problems. Firstly, a test was prepared in collaboration with J. Liljencrants at the Speech Transmission Laboratory of the Royal Institute of Technology in Stockholm and this time, the OVE III synthesizer was used. 42 test sentences 'byla to sese' - meaning 'it was/was it a session' - were prefabricated, in which the fundamental frequency values of the last vowel were systematically changed with much closer graduations than in the earlier test. The changes in the fundamental frequency corresponded to 0,  $\frac{1}{2}$ , 1, 2, 4, 6, and 8 semitones in each direction away from the fundamental frequency of the previous syllable; in addition, the intensity of the last vowel was changed at 3 levels. This material was then prepared for a new listening test with other groups of listeners: firstly, the isolated stimuli 'sese' were extracted from the recording and again the listeners' task was to mark the syllables they thought stressed; in the second test, the whole synthetic sentences were used and the listeners' task was to indicate with each item, whether they felt the sentence to be

statement, question or continuative sentences. A total of 100 listeners took part in the test, and the results are presented in the following graph:



Upper part of the graph: percentages of judgments 'stress on the second syllable'

Lower part of the graph: percentages of judgments S-statement, Q-question, C-continuative sentence

Scale in semitones: difference in pitch of the second syllable of the test word

It can be seen from the graph that the evaluation of stress accords well with the previous experiment: a steep rise from a dip of the curve in the middle with a slow decrease in percentages 'stressed' farther away from the center line. The graph also very clearly shows that in the area of fundamental frequency changes where the number of judgments 'stressed' begins to fall, there is a marked distinction of sentence melody - in the lower part of the graph - when it comes to evaluating the 'sentence' stimulus. By and large, these results corroborate the hypothesis that smaller changes of the fundamental frequency contribute more to the perception of stress, while greater intervals are more at play in the domain of sentence intonation.

In this test again the listeners responded readily and within narrow limits to signals of identical fundamental frequency; these stimuli with no or only a very small difference between the first and the second syllable were, however, not only identified, but the judgment 'first syllable stressed' was ascribed to them quite consistently. A possible interpretation of this finding is that the listeners evaluated both the rise and fall in frequency as a deviation from a pitch level of the first syllable in the test word, held constant throughout the whole test. This would be in agreement with findings from analyses of running speech, where departures from the basic contour of intonation in the direction up or down both are used in Czech to express prominence of a part of an utterance.

Two different tests were then prepared to investigate this hypothesis: in one test an attempt was made to keep the characteristics of the test similar to those of the previous experiment with the exception of the fundamental frequency of the first syllable of the test word: this was changed in a random order within an octave interval. The test items were prepared by means of the synthesizer HO 2, constructed by Maláč et al. (1975), and the listening tests were finished and evaluated in August 1978. In another test again the sequence 'sese' was used, but this time as a natural speech signal. In view of the high occurrence rate of 'se', even in iteration, in Czech texts it was easy to prepare a continuous text, a short story, which contained within its two pages a total of 116 repetitions of the syllable 'se' with the stress assumed to fall variously on the first or on the second

syllable of the sequence se-se. A professional speaker read the story and the tape recording was then used to prepare test tapes containing copies of all the sese combinations cut out of the master tape. A total of 50 listeners then heard the isolated items; their task was as with the earlier tests.

The results of the perception tests with both synthetic and natural items in isolation compare well with those of the earlier tests with synthetic stimuli, with a very clear difference, however, in the perception of changes of pitch: in the present experiments the effect of a change, i.e. the influence of pitch rise on an increase of judgments 'stressed', is manifest primarily towards the upper pitch levels. Clearly, this difference is not due to any difference inherent in the use of synthetic or natural speech signals, but to the different way in which the experiment was organized. In the first two experiments the fundamental frequency level of the first syllable of the test word was constant and hence probably provided a reference level comparable with a basic contour of sentence intonation. - In short, the results of the experiment with the isolated items of natural speech can be described as follows: the syllables marked 'stressed' had a higher fundamental frequency than those marked 'unstressed' in 93% of the cases, a higher peak intensity in 79%, and a longer duration in 44% of the cases.

#### Conclusions

By means of listening tests using disyllabic items, the influence of changes in intensity, fundamental frequency, and duration on the perception of stress in Czech can be shown.

In general, it can be demonstrated that an increase of any of these parameters leads to an increase of the number of judgments 'stressed'.

With changes of the fundamental frequency, however, the growth of the judgments 'stressed' is distinctly non-linear: slight changes of approximately a semitone lead to a considerable increase in the number of judgments 'stressed', while larger changes have a lesser effect on the evaluation of stress in a group of Czech listeners.

Results of a test with identical synthetic stimuli once in isolation and then in a simple sentence context corroborate the hypothesis that small changes of fundamental frequency have a

stronger effect on word-stress evaluations, whereas more marked changes have a noticeable effect on evaluations of sentence intonation.

The influence of context may be strong even in tests with isolated items: in tests with a constant pitch of the first syllable of disyllabic test items, deviations - both up and down - from this constant level are found to increase the number of syllables marked as stressed. In tests without the constant pitch level of the first syllable, only increase in fundamental frequency is found to add to the number of judgments 'stressed'.

Listening tests with isolated items only have been referred to in the present paper; it is obvious that even here there is a strong tendency of the listeners to evaluate the items as parts of a broader context.

#### Bibliography

- Bolinger, D.L. (1958): "On intensity as a qualitative improvement of pitch accent", *Lingua* 7, 175-182.
- Fry, D.B. (1958): "Experiments in the perception of stress", *L&S* 1, 126-152.
- Fyodorova, N. (1973): "The effect of some acoustic parameters of the synthetic speech signal on the perception of stress by Russian listeners", in *Speech Analysis and Synthesis III.*, W. Jassem (ed.), 229-248, Warsaw: PWN.
- Gårding, E. and A.S. Abramson (1965): "A study of the perception of some American English intonation contours", *SL* 19, 61-79.
- Hadding-Koch, K. and M. Studdert-Kennedy (1964): "An experimental study of some intonation contours", *Phonetica* 11, 175-178.
- Hála, B. (1967): *Výslovnost spisovné češtiny I.*, Praha: Academia.
- Janota, P. (1967): "An experiment concerning the perception of stress by Czech listeners", *Acta Universitatis Carolinae - Philologica* 6, *Phonetica Pragensia*, 45-68.
- Janota, P. and J. Ondráčková (1975): "Some experiments on the perception of prosodic features in Czech", in *Auditory Analysis and Perception of Speech*, G. Fant and M.A.A. Tatham (eds.), 485-496, London: Academic Press.
- Janota, P. and Z. Palková (1974): "The auditory evaluation of stress under the influence of context", *Acta Universitatis Carolinae - Philologica, Phonetica Pragensia IV.*, 29-59.
- Jassem, W., J. Morton and M. Steffen-Batóg (1968): "The perception of stress in synthetic speech-like stimuli by Polish listeners", in *Speech Analysis and Synthesis I.*, W. Jassem (ed.), 289-308, Warsaw: PWN.
- Lehiste, I. (1970): *Suprasegmentals*, Cambridge, Mass.: MIT Press.
- Lieberman, P. (1960): "Some acoustic correlates of word stress in American English", *JASA* 32, 451-454.
- Maláč, V., M. Ptáček, P. Dvořák and B. Borovičková (1975): "The HO 2 system - a digitally controlled terminal analog synthesizer", *Tesla Electronics*, 121-124.
- Mol, H. (1972): "The investigation of intonation", *Acta Universitatis Carolinae - Philologica* 1, *Phonetica Pragensia III.*, 176-178.
- Morton, J. and W. Jassem (1965): "Acoustic correlates of stress", *L&S* 8, 159-181.
- Ondráčková, J. (1961): "On the problem of the function of stress in Czech", *Zs.f.Ph.* 14, 45-54.
- Rigault, A. (1970): "L'accent dans deux langues à accent fixe: le français et le tchèque", in *Prosodic Feature Analysis*, P.R. Léon, G. Faure, A. Rigault (eds.), 1-12, Montréal: Didier.
- Rigault, A. and T. Arkwright (1972): "Les paramètres acoustiques de l'accent en tchèque", *Proc.Phon.* 7, 1004-1011.
- Romportl, M. (1973): "On the synonymy and homonymy of means of intonation", in *Studies in Phonetics*, M. Romportl, 137-146, Prague: Academia.
- Uldall, E.T. (1962): "Ambiguity: question or statement? or 'Are you asking me or telling me?'"', *Proc.Phon.* 4, 779-783.

## WORD STRESS AND SENTENCE STRESS IN VARIOUS TONE LANGUAGES

Eunice V. Pike, Summer Institute of Linguistics  
7500 West Camp Wisdom Road, Dallas, Texas 75236 USA

Introduction

The nine languages summarized here all use two or more tones as part of the features which contrast lexical items. All but one of the languages, Fasu (4), show tonal contrasts on both stressed and nonstressed syllables. In Fasu the tonal contrasts occur on stressed syllables only. One language, Mikasuki (6), contrasts long versus short vowels in addition to tone.

Instead of contrasting lexical tone in the prepause syllable, Tenango Otomi (3) has a different pitch system on that syllable, one that indicates the attitude of the speaker.

Diuxi Mixtec (9) has two types of word stress, one marked by vowel length, and the other by allotones. Some words have both types, but others have only the type marked by vowel length. Eastern Popoloca (2) also has two types of word stress, one marked by vowel length and the other by consonant length. In this language the two types never occur in the same word.

Probably in all of the languages, loudness is optionally present on stressed syllables. Vowel length is used to mark stress in at least five of the languages. Consonant length is one of the features marking stress in at least three languages.

By the term "word stress", I mean the syllable which is the nucleus of a rhythm wave, in this case, the phonological word (Pike 1976, 54-69). By "sentence stress" I mean the nucleus of a larger rhythm wave. As I have used it in this paper, this rhythm wave coincides with the pause group (K. Pike 1967, 392-403), so by "sentence stress" I mean the syllable which is the nucleus of a pause group.

A word uttered in isolation is between pauses; therefore the stressed syllable of a word in isolation has sentence stress. To identify word stress, usually there must be at least two words in the utterance. Word stress is the stress which remains on a word when it occurs in the margin of a pause group.

If, in a specific language, word stress and sentence stress occur on the same syllable, it is perhaps impossible to know which features are marking word stress when the words are studied only in isolation. It may be, for example, that in Mikasuki (6) some of the features which were described as those of word stress were actually features of sentence stress, since the data, for the most part, were studied from words uttered in isolation.

There is less apt to be confusion when sentence stress occurs on a different syllable from that where word stress occurs. For example, in Ayutla (7) and Acatlan Mixtec (8), and also in Tenango Otomi (3) word stress occurs on the first syllable of the stem, but sentence stress (with some exceptions) occurs on the prepause syllable.

Perhaps the thing that surprised me most, as I summarized the nine languages, was that in only one of the languages, Fasu (4), were vowel allophones spoken of as being determined by their occurrence in relation to a stressed syllable. If I have the opportunity to hear these languages again, that is one of the points I will check.

Languages summarized

(1) Marinahua of Peru (Pike and Scott 1962) has a contrastive tone, high versus low, on each syllable (p.7). Word stress occurs on the first syllable of the stem and is marked by vowel length (p.4). Sentence stress occurs on the first syllable of the stem

of the last word in the sentence (p.2); it has an even longer vowel than occurs with word stress. Sentence stress is also marked by allotones, that is, the high is raised, and the low tone usually glides down when in a syllable with sentence stress (p.8). When the speaker is irritated, the last consonant of the sentence may be lengthened, and the loudness may shift from the normally stressed syllable to the prepause syllable (p.4).

(2) In Eastern Popoloca of Mexico (Kalstrom and Pike 1968), contrastive tone consists of four level tones plus ten tone clusters which are combinations of those tones. There are two types of stress. Some words have stress marked by a long vowel (that stress occurs on the next to the last syllable of the stem, p.16,18). Other words have a stress that is marked by a long consonant (that stress usually occurs on the last syllable of the stem (p.17). When only one consonant occurs in the syllable, that consonant is lengthened. When a consonant cluster occurs, it is the /h/, /?/, or /n/ of the cluster which is lengthened--all consonant clusters have either /h/, /?/, or /n/. All words have either one stress type or the other, but no words have both.

In a normal sentence, the prepause syllable is usually short, louder than other syllables in the sentence, and it ends in a sharp glottal stop (p.28). A polite sentence is raised in key and the prepause syllable is long, lenis, gradually getting softer as it glides upward. There is no final glottal stop in that type of sentence (p.29).

(3) In Tenango Otomi of Mexico (Blight and Pike 1976), high, low and upgliding tone contrast lexical items. The upgliding tone occurs only on syllables with word stress. Word stress occurs on the first syllable of the stem, and is marked by vowel length

(p.56); it is also marked in that voiceless stops are preaspirated when they are initial in a stressed syllable (p.52). A low tone in a stressed syllable is slightly lower than a nonstressed low; a stressed high usually has a slight downgliding allotone when occurring between voiced consonants (p.55). Sentence stress occurs on a prepause syllable and is marked by loudness. It is that prepause syllable that carries contrastive intonation. There is no contrast of lexical tone on the prepause syllable nor on any word-final syllable (p.55).

(4) In Fasu of Papua New Guinea (May and Loeweke 1965), a contrast of high versus low tone occurs on only one syllable per word--the stressed syllable; the placement of that syllable is not predictable. Vowels /i/ and /e/ have open variants which fluctuate with close variants when prestressed if the stressed vowel is /i/ or /e/ respectively (p.92). Within a sentence, there is a gradual downdrift of pitch. In a question-doubt sentence, without an interrogative marker, there is a small upglide on the final syllable (p.95). Other attitudes of the speaker may be indicated by a wider spread in the tone levels, or by voice quality, etc.

(5) In Golin of Papua New Guinea (Bunn 1970), there is contrastive tone, high versus low, on each syllable. Word stress occurs on the final high, or if there is no high, then on the final low (p.4). Sentence stress occurs on the same syllable as word stress, but it is louder and if it is a syllable with high tone, it usually has a higher allotone. When sentence stress occurs on a syllable with low tone, there is optionally a lower allotone (p.5). Sentence stress may occur on any word of the sentence; it is usually used for emphasis (p.6).



(6) Mikasuki of Florida USA (West 1962) has contrastive long versus short vowels. Mikasuki also has three levels of tone and one tone cluster which are used in lexical items (p.82). In a question, one of the syllables of the sentence may have an extra-high tone and additional length, and one of the words may end in glottal stop (p.82,89). (Glottal stop is not lexically pertinent.) Other clause types, negative statements, imperatives, for example, may also have the extra-high tone or tone clusters that do not occur in simple lexical items (p.89). Word stress usually occurs on the highest nonfinal syllable (p.85). In the normal sentence, there is a gradual drop of pitch between words (p.88-89).

(7) In Ayutla Mixtec of Mexico (Pankratz and Pike 1967), there are three levels of tone, contrastive on each syllable (p.291). Word stress occurs in the first syllable of the stem and is marked by loudness and also by allophones of the consonants. That is, when contiguously following the stressed syllable, voiceless stops and affricates are preaspirated, voiced continuants are lengthened, voiceless continuants are either lengthened or preceded by a slight hiatus (p.288). Allotones mark word stress in that a proclitic with high tone is not as high as a stressed high which immediately follows it (p.291). Throughout the sentence there is a downdrift of pitch. Sentence stress is usually louder than other syllables, and occurs either on the prepause syllable, or on the syllable with word stress-- there is variation in accordance with the CV pattern and the tone sequence of the last word (p.294).

(8) Acatlan Mixtec (Pike and Wistrand 1974) has contrastive tone on each syllable: low, mid, high and up-step (p.83). Word

stress occurs on the first syllable of the stem. It is marked by allophones of consonants (p.100,103) in that all consonants, except the flapped /r/, are lengthened when contiguously following a stressed syllable. If the syllable which follows stress does not begin with a consonant, then it is the stressed vowel which is lengthened. Sentence stress occurs on the syllable immediately preceding pause (p.104). There is general downdrift in relaxed speech in that each low tone tends to be lower than the preceding low (p.84).

(9) Diuxi Mixtec (Pike and Oram 1976) has two contrastive tones, high versus low, one of which occurs on each syllable. There are two types of word stress. The type marked by a lengthened vowel occurs, on each word, on the first syllable of the stem (p.322). The second type of word stress occurs on the word-final syllable, but only on some words. It is marked by allotones. That is, a stressed high between another high and pause has a sharp downglide. When between low and pause, the high may not downglide, but it is definitely higher than a nonstressed high in that environment. A stressed low tone downglides from a point starting noticeably higher than a nonstressed low (p.325).

References

- Blight, Richard C. and Eunice V. Pike (1976): "The Phonology of Tenango Otomi", IJAL 42, 51-57.
- Bunn, Gordon and Ruth (1970): "Golin Phonology", (Papers in New Guinea Linguistics XI) Pacific Linguistics A 23, 1-7.
- Kalstrom, Marjorie and Eunice V. Pike (1968): "Stress in the Phonological System of Eastern Popoloca", Phonetica 18, 16-30.
- May, Jean and Eunice Loeweke (1965): "The Phonological Hierarchy in Fasú", Anthropological Linguistics 7.5, 89-97.
- Pankratz, Leo and Eunice V. Pike (1967): "Phonology and Morphotonemics of Ayutla Mixtec", IJAL 33, 287-99.
- Pike, Eunice V. (1976): "Phonology", in Tagmemics I, Aspects of the Field, Ruth M. Brend and Kenneth L. Pike (eds.), 45-83.
- Pike, Eunice V. and Joy Oram (1976): "Stress and Tone in the Phonology of Diuxi Mixtec", Phonetica 33, 321-33.
- Pike, Eunice V. and Eugene Scott (1962): "The Phonological Hierarchy of Marinahua", Phonetica 8, 1-8.
- Pike, Eunice V. and Kent Wistrand (1974): "Step-up Terrace Tone in Acatlán Mixtec (Mexico)", in Advances in Tagmemics, Ruth M. Brend (ed.), 81-104, Amsterdam: North Holland Pub. Co.
- Pike, Kenneth L. (1967): Language in Relation to a Unified Theory of the Structure of Human Behavior, (2nd revised ed.) The Hague: Mouton.
- West, John David (1962): "The Phonology of Mikasuki", Studies in Linguistics 16, 77-91.

LEXICAL STRESS, EMPHASIS FOR CONTRAST, AND SENTENCE INTONATION  
IN ADVANCED STANDARD COPENHAGEN DANISH

Nina Thorsen, Institute of Phonetics, University of Copenhagen

Due to lack of space, no references will be made to the very considerable literature on intonation in other languages, nor will any extensive documentation be given.

1. A model of Danish intonation

Intonation in short sentences in Advanced Standard Copenhagen (ASC) Danish may be presented as in fig. 1, which is but a model, with the advantages and shortcomings that modeling almost always entails in terms of simplicity and inaccuracy, respectively. It is based on recordings by six subjects, three males and three females, of a rather elaborate material (Thorsen 1978a, 1979b). The qualitative statements which can be read off the figure are perfectly representative of all the subjects, but the quantifications involved are, of course, averages, and no one subject behaves as mathematically neatly as the model would have you believe.

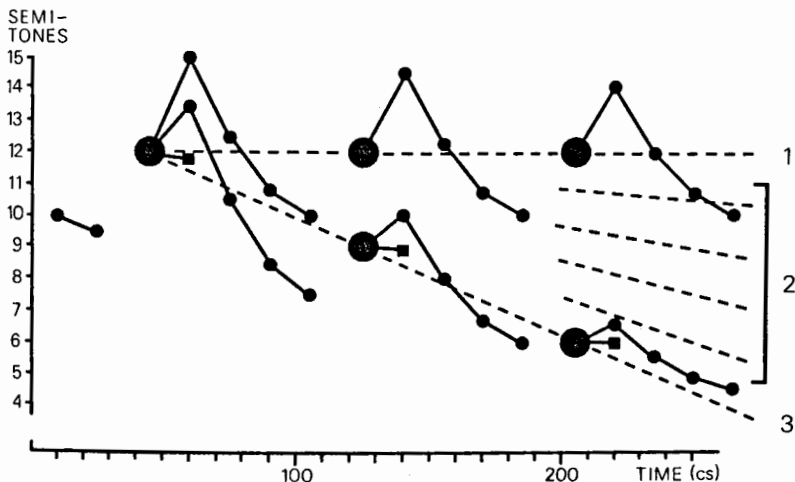


Figure 1

A model for the course of  $F_0$  in short sentences in ASC Danish. 1: statement questions, 2: interrogative sentences with word order inversion and/or interrogative particle, and non-final periods (variable), 3: declarative sentences. The large dots represent stressed syllables, the small dots unstressed ones, and the small squares represent an unstressed syllable being the only one between two stressed ones (see further the text). The full lines represent the  $F_0$  pattern associated with stress groups, and the broken lines denote the intonation contours.

A basic assumption underlying fig. 1 is that the complex course of fundamental frequency (Fo) in an utterance is the outcome of a superposition of several components: (1) A sentence component which supplies the INTONATION CONTOUR (broken lines). (2) On the contour is superposed a stress group component which furnishes the STRESS GROUP PATTERNS (full lines). (3) To the resultant of those two components is added, in words containing stød, a stød component, rendering STØD MOVEMENTS. However, as stød words had been excluded from the material, the model does not include this particular feature. These first three components are language specific and thus "speaker controlled". (4) Finally, intrinsic Fo level differences between segments, and coarticulatory variations at segment boundaries supply a MICROPROSODIC COMPONENT, which is not consciously controlled by the speaker, but due to inherent properties of the speech production apparatus and which, therefore, is superfluous in the model from the point of view of the human speaker. - This concept of "layers" in intonation is anything but original; the triviality of the statement does not, however, deprive it of its validity or its relevance: firstly, it is tremendously useful in the interpretation of Fo tracings (Thorsen forthcoming) and, secondly, it has a very direct bearing on the theme of this symposium, 'The relation between sentence prosody and word prosody (stress and tone)':

The relation between word stress and sentence prosody (i.e. sentence intonation: duration and intensity are not considered) is physically a very close-knit and intricate one, but on the higher and more abstract levels we may hypothesize that very little of the mutual influence which is customary in a relationship takes place, as long as we are dealing with neutral lexical stresses.

#### 1.1 Stress group patterns

As can be seen from the full lines of fig. 1, the stress group pattern can be described as a (relatively) low stressed syllable followed by a high-falling tail of unstressed syllables. This pattern is a predictable and recurrent entity, though allowing for contextual variation in the magnitude of the rise from stressed to unstressed syllable and in the slope of the fall through the unstressed ones. It was this observation which gave rise to the definition of the STRESS GROUP in ASC Danish as A STRESSED SYLLABLE PLUS ALL SUCCEEDING UNSTRESSED SYLLABLES (within the same, non-compound sentence), irrespective of intervening word or morpheme boundaries, and, as a consequence of this predictability and recurrency, it al-

so brought about the definition of the INTONATION CONTOUR as THE COURSE DESCRIBED BY THE STRESSED SYLLABLES ALONE.

#### 1.2 Intonation contours

The intonation contours tend to vary systematically with sentence type, declarative sentences having the most steeply falling contours, at one extreme, and statement questions (i.e. questions with statement syntax where only the intonation contour signals their interrogative function) having "flat" contours, at the other extreme. In between these two are found other types of questions as well as non-final periods. For a further account of these contours and their perception, see Thorsen 1978b, 1979a.

#### 2. Implications of the model

##### 2.1 Fo movements in syllables

The model does not specify the Fo movements of syllables: the tonal composition of the stress group pattern as one of LOW plus HIGH FALLING allows for a very simple account of Fo movements in vowels and consonants: segments do not carry specific movements (except when stød is involved) but simply float on the Fo pattern, and slight variations in Fo movement would be due then to the fact that segments do not always hit the patterns at exactly the same place.

##### 2.2 The course of the intonation contour

(a) When the number of stress groups changes, everything else being equal, so does the slope of a given contour, leaving only the "flat" ones intact; the constancy presumably lies in the interval between the first and the last stressed syllable, with intervening stressed syllables evenly distributed between them, and not in a certain rate of change (this point needs further verification which I hope to present orally in August).

(b) When the number of unstressed syllables varies in the stress groups, the stressed syllables will not be equidistantly spaced in time, and the straight lines of fig. 1 break up into a succession of shorter ones with unequal slopes.

Combining the effects of changes of both types leaves us with an infinity of physically different intonation contour configurations. On a higher level in production these variations may not exist, and perceptually they may be obliterated, turning the contours into smoothly slanting slopes, (1) if what we aim at producing and what we perceive are equal intervals between stressed syllables and not the actual slope of the contour, and (2) if we assume that isochrony, be it not a physical reality, is a psycho-

logical reality with the speaker/listener.

### 2.3 Fo patterns of stress groups

#### 2.3.1 Stress groups with more than one unstressed syllable

(a) In statements, the rise from stressed to unstressed syllable is, on the average  $1\frac{1}{2}$ , 1, and  $\frac{1}{2}$  semitone, respectively, in the first, second, and third stress group. In statement questions, the rises amount to 3,  $2\frac{1}{2}$ , and 2 semitones, respectively. The difference in magnitude of this rise, between patterns riding on different contours is very likely a direct consequence of differences in the level of the following stressed syllable.

(b) The decrease with time in the rise from stressed to unstressed syllable is the same in statement questions and statements, one semitone. This decrease, which is independent of the particular contour, may be seen as a consequence of either of two distinct processes, or of a combination of them. It may be a "voluntary" decrease, i.e. a signal of finality, and/or it may be a physiological phenomenon: the closer you get to the end of the utterance, the less "energy" is expended and the less complete the gestures will be; either or both phenomena may also account for the less and less steep falls through the unstressed syllables.

If the variation in the Fo patterns with intonation contour and time is physiologically determined, the speaker may be unconscious of it, and the listener may neglect or compensate for it.

#### 2.3.2 Stress groups with only one unstressed syllable

Stress groups with one unstressed syllable will of course be shorter than those with several, a feature which is not reflected in fig. 1. - A single unstressed syllable does not accomplish a full rise-fall when the following stressed syllable is considerably lower than the preceding one, as is the case in statements. Instead it lands on very nearly the same level or slightly below the preceding stressed syllable and, accordingly, the rise-fall is amputated. A full rise-fall may be intended by the speaker and the amputation be due to a shortcoming in the peripheral speech production mechanism. Accordingly, the listener may well re-introduce a rise-fall (this is, indeed, my own subjective impression). But we have here an indication that time (rhythm) overrides Fo when the two are in conflict. On the other hand, there is definitely a tendency towards as complete rise-falls as possible. Two unstressed syllables will traverse more than half the fall exhibited by four, everything else being equal. - These two facts together are

another reminder that speech is not a card-board structure but a smooth and dynamic process.

### 2.4 Conclusion

If the assumptions made about production and perception of Fo courses hold water, we are left with two components which physically are highly interactive but on more abstract levels may be invariant, apart from the fact that contours change with sentence type.

#### 3. Emphasis for contrast

By emphasis for contrast is meant the extra prominence on one of the syllables in the utterance, used to denote a contrast which may be implicit or may be explicitly stated in the context. I have deliberately avoided terms like 'focus', 'sentence accent', or 'nucleus' because these terms are used, in a number of languages, to describe a phenomenon different from emphasis for contrast: one of the lexically stressed syllables in the utterance will always have slightly greater prominence (realized, very roughly speaking, as a more elaborate Fo movement within that syllable), and if nothing else is specified by the context, it will fall on the last stressed syllable. - A similar phenomenon does not exist in ASC Danish as a thing apart from emphasis for contrast. Whenever and wherever such a slightly heavier stress is introduced, it invariably invokes the impression of contrast. Insofar as we are not faced with incomplete evidence or with a false dichotomy, i.e. one due to differences in concepts, Danish seems to be markedly different from e.g. English, German, and Swedish.

#### 3.1 Contrastive stress and Fo

The following account is based on a material of sentences, uttered in dialogues, where the contrasts were all explicitly stated in the context (but I strongly believe that they would have looked no different had they been implicit). - When emphasis for contrast occurs, it affects the intonation contour as well as Fo patterns.

Fig. 2 compares the neutral edition with three statements where the emphasis lay on the first, second, and third lexically stressed syllable. (Durational differences between the neutral and emphatic editions are very slight and there is no doubt that Fo is the prime cue to contrast, as it is to neutral lexical stress.) The obvious changes introduced in the Fo course by emphasis for contrast is a raising of the syllable in question (represented by a star), a drastic fall from first to second unstressed syllable, plus a not inconsiderable shrinking of the surrounding Fo patterns:

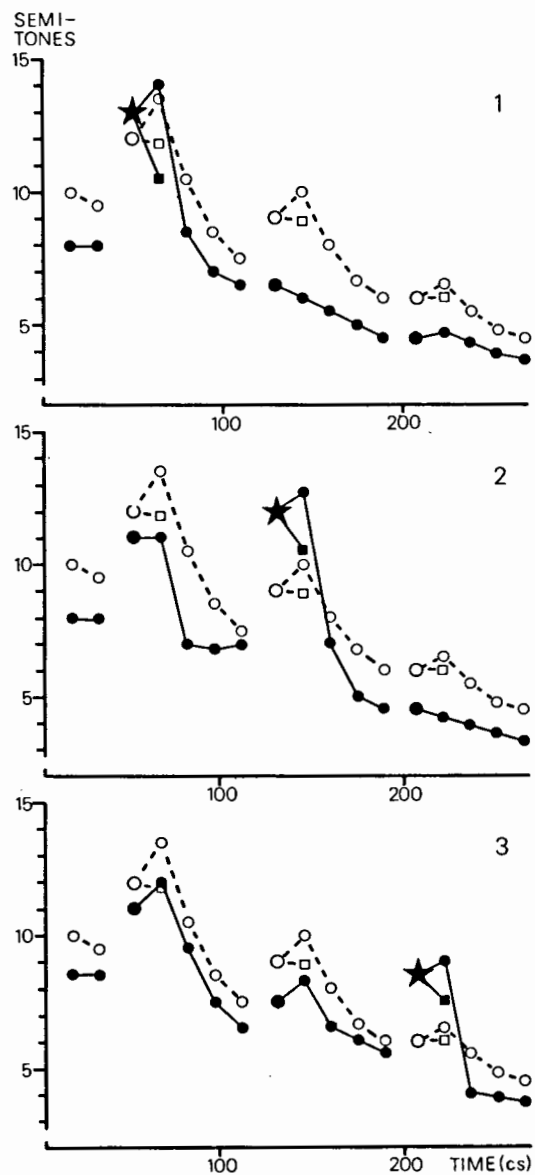


Figure 2

Models for statements with emphasis for contrast on (1) the first, (2) the second, and (3) the last stressed syllable, compared to the neutral edition (broken lines and empty dots).

(1) when emphasis is on the first stressed syllable, it is higher, the rise to the first unstressed syllable is smaller, and the fall through the following unstressed syllables is steeper than for the neutral case. The levels of the second and third stressed syllables are considerably lowered and the LOW+HIGH FALLING pattern is annihilated in the second and shrunk in the third stress group. The syllables of the second and third stress groups look, tonally, more like a series of unstressed syllables continuing the fall in the first one.

(2): emphasis on the second stressed syllable repeats the pattern of (1), and we also get a certain reduction of the first stress group with a very steep fall from first to second unstressed syllable.

(3) the pattern repeats itself in the last stress group, with a shrinking of the preceding ones as well.

Again we note that a single unstressed syllable does not accomplish a full rise-fall but instead drops well below the preceding emphatic one.

The feature common to

the three cases seems to be that the syllable on which the emphasis for contrast occurs must stand out clearly from the surroundings, which is brought about by a raising of that syllable as well as by a lowering of the immediate surroundings, except for the first of several post-tonic syllables. The change is slightly greater in the succeeding than in the preceding Fo course. During some informal experiments performed with the ILS-system for analysis and synthesis at the Institute of Linguistics, Uppsala University, it appeared that shrinking the Fo course in the surroundings is sufficient to create the impression of emphasis for contrast. To get emphasis on the word 'sidste' in the statement 'Det er sidste bus til Tiflis.' (It's the last bus for Tiflis.) it is sufficient to change the rise from 'bus' to 'til' to a level or a slight fall, whereas just raising the stressed syllable of 'sidste' will not do the trick. Likewise, to get emphasis on 'bus', lowering the unstressed syllable of 'sidste' will do and just raising 'bus' does not accomplish anything.

The three Fo courses in fig. 2 look widely different and only vaguely resemble fig. 1 "3" although the utterances still sound declarative. What constitutes the intonation contour in utterances with emphasis for contrast, I hesitate to say at present. They may resemble one-word utterances in that the difference between statement and question lies in the level of and movement within the emphatically stressed syllable as well as in the course of the succeeding unstressed ones (Thorsen, 1978a), or the intonation contour may be extrapolated from, and thus still be definable in terms of, the lexically stressed syllables surrounding the emphatic one. The first solution would be interesting, because it implies that in utterances with emphasis, word prosody takes precedence over sentence prosody, whereas the second solution would make the definition of intonation contour apply to a wider range of utterances.

- References (ARIPUC = Ann. Rep. Inst. Phonetics, Univ. Copenhagen)
- Thorsen, N. (1978a): "An acoustical investigation of Danish intonation", *JPh* 6, 151-175.
  - (1978b): "On the identification of selected Danish intonation contours", *ARIPUC* 12, 17-73.
  - (1979a): "Aspects of Danish intonation", *Travaux de l'Institut de Linguistique de Lund* 13 (in print).
  - (1979b): "Lexical stress, emphasis for contrast, and sentence intonation", *ARIPUC* 13 (in preparation).
  - (forthcoming): "Interpreting raw fundamental frequency tracings of Danish", *Phonetica*.

## SENTENCE TONE IN SOME SOUTHERN NIGERIAN LANGUAGES

Kay Williamson, School of Humanities, University of Port Harcourt, Port Harcourt, Nigeria

The lexical tones of words can be modified in various ways:

1. By essentially phonetic rules, such as tone-spreading (Hyman & Schuh, 1974): e.g. Yoruba /Low-High/ → [Low-Rising] because Low 'spreads' into the following High. Such phonetic rules result in phonological change if the conditioning factor is lost.
2. By morphophonemic rules, i.e. rules whose phonetic motivation is no longer obvious: e.g. in the Kolokuma dialect of Izon two words which have the same tone pattern in isolation may have different tonal effects upon the following word (cf. Williamson, 1965).
3. By the interaction of purely tonal morphemes with the tones of the normal morphemes which consist of both segments and tones: e.g. the subject concord marker of Edo is analysed by Amayo (1975) as having lost its segmental features in practically all contexts, so that its presence is normally detected only by its tonal effects on neighbouring morphemes. Purely tonal morphemes appear to be restricted to common grammatical elements.

Lexical tones are also modified to show sentence type. In some languages such modifications involve changing the absolute but not the relative pitch of sentences: e.g. in Kana (Ogoni group: tonemes High, Mid, Low):

Statement: Lo tɔ. [--] 'The house.'  
 Question: Lo tɔ? [^^] 'The house?'  
 Exclamation: Lo tɔ! [^^] 'The house!'

The basic Mid-Mid tone (seen in the statement) is raised for a question and raised even more for an exclamation (this is indicated phonetically by writing it above the square brackets, i.e. outside the normal voice range). This type of modification is here called intonational, and is regarded as comparable to what obtains in a non-tone language.

Other languages have a second type of modification co-existing with the intonational type. This involves a change of the tone pattern, not simply a general modification of the absolute pitch, and is here called sentence tone. In the examples that follow, the tone system of each language will be summarized and then the sentence tone modifications will be stated.

A. YEKHEE (=ETSAKO), Ekpheli dialect (North-Central Edoïd group), Elimelech (1976):

Basic tones: H, L

Tone rules: a) downdrift on each series of highs separated by low  
 b) falling and rising glides formed from HL, LH  
 c) downstep from simplification of rising glide

Sentence tone: (for nouns; verbal sentence questioning said to be different but not specified, Elimelech, 1976, 50):

1. statement: additional final low added to final high
2. question: additional final high added to statement tone pattern

Lexical tone patterns of disyllabic nouns: LL, HL, HH

Data:	Statement	Question
1. LL 'cup'	Àkpà. [--]	Àkpà? [^^]
2. HL 'house'	Ówà. [^^]	Ówà? [^^]
3. HH 'axe'	Údzé. [^^]	Údzé? [^^]

B. DEGEMA (Delta Edoïd group), personal investigation, analysis tentative:

Basic tones: H, L; downstep probably predictable

Tone rules: a) downdrift on each series of highs separated by low  
 b) falling glide formed from HL  
 c) the final low in a series of lows becomes high (under certain conditions)  
 d) all but the first of a final series of highs are downstepped

Sentence tone: 1. statement: basic tones + tone rules  
 2. question: final low added to statement tone pattern, combines variously with preceding tone  
 3. exclamation: general raising of statement tone

Lexical tone patterns of disyllabic nouns: LL, HH, HL (loanwords only):

Data	Statement	Question	Exclamation
1. LL 'head'	Útóm. [^^]	Útóm? [^^]	Útóm! [^^]
2. HH 'river'	Édá. [^^]	Édá? [^^]	Édá! [^^]
3. HL 'cat'	Pòsì. [^^]	Pòsì? [^^]	Pòsì! [^^]
4. LL (?)	Mòyá. [^^]	Mòyá? [^^]	Mòyá! [^^]
	'He is coming.'	'Is he coming?'	'He is coming!'
5. HH (?)	Àbò. [^^]	Àbò? [^^]	Àbò! [^^]
	'He is there.'	'Is he there?'	'He is there!'

## C. ISOKO (Southwestern Edoïd group), Elugbe (1977):

Basic tones: H, LTone rules: a) falling glide formed from HL  
b) no downdriftSentence tone: 1. statement: final series of lows raised to mid  
2. question: additional final low added  
3. exclamation: no raising of final series of lowsLexical tone patterns of disyllabic nouns: LL, HH, HL

Data:	Statement	Question
1. LL 'native doctor'	Ọ̀bù. [--]	Ọ̀bù? [ _ ]
2. HH 'warrior'	Ọ̀gbá. [ ^ ]	Ọ̀gbá? [ ^ ]
3. HL 'maize'	Ọ̀kà. [ ^ ]	Ọ̀kà? [ ^ ]

## D. IZON, Kolukuma dialect (Ijọ group), personal investigation:

Basic tones: H, LTone rules: a) downdrift on each series of highs separated by low  
b) complex morphophonemic rulesSentence tone: 1. statement: basic tones + tone rules  
2. question: slight raising of highs, cancellation of downdrift, final low added  
3. exclamation: general raising of highs and of final low; cancellation of downdrift  
4. command: slight raising of highs, cancellation of downdriftLexical tone patterns of disyllabic nouns: LH (3 types), HH (2 types), HL

Data:	Statement	Question	Exclamation
1. LH 'yam'	Bùrú. [ _ ]	Bùrú? [ _ ]	Bùrú! [ _ ]
2. HH 'medicine'	Dírí. [ ^ ]	Dírí? [ ^ ]	Dírí! [ ^ ]
3. HL 'sail'	Bálà. [ ^ ]	Bálà? [ ^ ]	Bálà! [ ^ ]
Statement	Question	Exclamation	Command
4. Wónì múdọ̀.	Wónì múdọ̀?	Wónì múdọ̀!	Wómìnì mú!
[ ^ _ - - ]	[ ^ _ - ]	[ ^ _ - ]	[ ^ _ - ]
'We have gone.'	'Have we gone?'	'We've gone!'	'Let's go!'

## E. NEMBE (Ijọ group), personal investigation:

Basic tones: H, LTone rules: a) downdrift on each successive high even without intervening low  
b) complex morphophonemic rulesSentence tone: 1. statement: final low tone becomes high  
2. question: final high tone becomes low  
3. exclamation: general raising of highs; cancellation of downdrift after low  
4. command: additional final low added to statement patternLexical tone patterns of disyllabic nouns: LH, LL, HL

Data:	Statement	Question	Exclamation
1. LH 'yam'	Bùrú. [ _ ]	Bùrú? [ _ ]	Bùrú! [ _ ]
2. LL 'book'	Dírí. [ _ ]	Dírí? [ _ ]	Dírí! [ _ ]
3. HL 'Ebi'	Ébí. [ ^ ]	Ébí? [ ^ ]	Ébí! [ ^ ]
Statement	Question	Exclamation	Command
4. Ébí ọ̀.	Ébí ọ̀?	Ébí ọ̀!	Ébí, ọ̀!
[ ^ _ - ]	[ ^ _ - ]	[ ^ _ - ]	[ ^ _ - ]
'Ebi came.'	'Did Ebi come?'	'Ebi came!'	'Ebi, come!'

## F. KALABARI, didlect of Eastern Ijọ, Jenewari (1977) and personal investigation:

Basic tones: H, L, distinctive downstep (´)Tone rules: a) downdrift on each series of highs separated by low  
b) complex morphophonemic rulesSentence tone: 1. statement: a) basic tones + tone rules (for non-emphasized nouns)  
b) basic tones + (H)´H (first H only after L)

(for verb forms ending H, NPs ending in pronoun/article, and emphasized nouns, especially in answer to a question)

2. question: basic tones + tone rules  
3. exclamation: as for 1b), plus general raising  
4. command: basic tones + tone rules + additional LLexical tone patterns of disyllabic nouns: LL, HH, H´H, HL, LH

Data:	Statement a) + Question	Statement b)
1. LL 'yam'	Bùrú. Bùrú? [ _ ]	Bùrú´. [ _ ]
2. HH 'book'	Dírí. Dírí? [ ^ ]	Dírí´i. [ ^ ]
3. H´H 'house'	Wá´rí. Wá´rí? [ ^ ]	Wá´rí´i. [ ^ ]
4. HL 'leopard'	Sírí. Sírí? [ ^ ]	Sírí´i. [ ^ ]
5. LH 'Gogo'	Gógó. Gógó? [ _ ]	Gógó´o. [ _ ]



	<u>Statement b)</u>	<u>Question</u>	<u>Exclamation</u>	<u>Command</u>
6.	ò ɓòtẹ́'ẹ́	ò ɓòtẹ́?	ò ɓòtẹ́'ẹ́!	ò ɓòo!
	[ _ _ \ ]	[ _ - - ]	[ _ - \ ]	[ _ \ ]
	'He has come.'	'Has he come?'	'He has come!'	'Let him come!'

G. IGBO, Green and Igwe (1963) and personal investigation:

Basic tones: H, L, distinctive downstep (')

Tone rules: a) downdrift on each series of highs separated by low  
b) falling and rising glides formed from HL, LH  
c) morphophonemic rules

Sentence tone: 1. statement: basic tones + tone rules  
2. question: a) intonational raising of high, etc., in nominal sentences  
b) inseparable subject pronouns change from H to L, in verbal sentences  
3. exclamation: a) intonational raising of high and lowering of low  
b) cancellation of downdrift  
4. command: basic tones + tone rules

Lexical tone patterns of disyllabic nouns: HH, LH, HL, LL, H'H

<u>Data:</u>	<u>Statement</u>	<u>Question</u>	<u>Exclamation</u>
1. HH 'head'	ísí. [ -- ]	ísí? [ -- ]	ísí! [ -- ]
2. LH 'rat'	òké. [ - - ]	òké? [ - - ]	òké! [ - - ]
3. HL 'house'	ùlọ. [ - _ ]	ùlọ? [ - _ ]	ùlọ! [ - _ ]
4. LL 'earth'	àlà. [ - - ]	àlà? [ - - ]	àlà! [ - - ]
5. H'H 'tooth'	é'zé. [ - - ]	é'zé? [ - - ]	é'zé! [ - - ]
6.	ó jèrè ahíá. [ - - - ]	ó jèrè ahíá? [ - - - ]	ó jèrè ahíá! [ - - - ]
	'He went to market.'	'Did he go to market?'	'He went to market!'

Summary and conclusions

The statement is the most basic form of sentence, for:

- a) Most commonly the statement form uses only basic forms plus general tone rules (Kana, Degema, Iẗon, Kalabari (type a), Igbo), whereas questions and exclamations usually require extra rules.  
b) Questions are normally formed by addition to the statement form (Yekhee), or to the basic/statement form (Degema, Isoko, Iẗon).  
c) The only case where the statement form appears more complex than the question is in Kalabari b), and this is apparently an emphatic

form which has become generalized for certain grammatical categories.

- d) Exclamations are formed by modification of the statement, never of the question.

CONCLUSION 1. The statement is the most basic sentence type.

This is probably a universal.

Questions are formed from statements by:

- a) addition of a floating low tone (Degema, Isoko, Iẗon)  
b) replacement of a high by a low tone (Nembe, Igbo in verbal sentences)  
c) addition of a floating high tone (Yekhee)  
d) general raising (Kana, Iẗon, Igbo in nominal sentences)  
e) cancellation of downdrift (Iẗon)

CONCLUSION 2. Yes/no questions are marked by a question marker, by intonational raising (including cancellation of downdrift), or by both. This is probably a universal.

Question markers are morphemes which are segmental + tonal or purely tonal (cf. the introduction). The floating low tones of Degema, Isoko, and Iẗon result from morphemes which have lost their segmental features; e.g. in Engenni (Delta Edoid group, closely related to Degema) the sentence-final question marker is à (Thomas, 1969).

It is also possible for the floating tone to replace the adjacent segmental tone (Nembe, Igbo). Historically, the floating tone first combines with the adjacent tone to form a glide (Yekhee, Degema, Isoko, Iẗon); later, the glide is simplified to a level tone:

Nembe: \*H + ' > \*F > L  
Igbo: \*' + H > \*R > L

CONCLUSION 3. Question markers are morphemes which in a tone language always include tone and sometimes lose their segmental features, after which they are realized like other floating tones. There is no universal that question markers must have high tone, or that they must be sentence-final.

Exclamations are marked by the raising of high tones (sometimes also of low tones) to a greater extent than in questions, and sometimes by the cancellation of downdrift (Iẗon, Nembe, Igbo) and the lowering of low tones (Igbo).

CONCLUSION 4. Exclamations are marked by the raising of tones, especially high ones, and by the increasing of the intervals between tones. This is probably a universal.

Commands seem not to be primarily marked by tone/intonation changes. In the reported cases they either have the same pattern as statements (Igbo) or the statement pattern with an additional floating tone marker (Nembe, Kalabari), or with slight raising and cancellation of downdrift (Izon).

CONCLUSION 5. Commands either resemble statements or differ from them only by the addition of an imperative marker, or by slight raising and elimination of downdrift. This is probably a universal.

References

- Amayo, A. (1975): "The structure of verbal constructions in Edo (Bini)", J. of West Afr. Lang. 10:1, 5-27.
- Elimelech, B. (1976): A Tonal Grammar of Etsako, UCLA Working Papers in Phonetics, 35.
- Elugbe, B.O. (1977): "Some implications of low tone raising in Southwestern Edo", Stud. in Afr. Ling., Supplement 7, 53-62.
- Green, M.M. and G.E. Igwe (1963): A Descriptive Grammar of Igbo, Berlin: Akademie-Verlag and London: Oxford University Press.
- Hyman, L.M. and R.G. Schuh (1974): "Universals of tone rules: evidence from West Africa", Linguistic Inquiry 5, 81-115.
- Jenewari, C.E.W. (1977): Studies in Kalabari Syntax, Ph.D. thesis, University of Ibadan.
- Thomas, E. (1969): A Grammatical Description of the Engenni Language, Ph.D. thesis, University of London.
- Williamson, K. (1965): A Grammar of the Kolokuma Dialect of Ijo, London: Cambridge University Press.

## PERCEPTION OF SPEECH VERSUS NON-SPEECH

## Summary of Moderator's Introduction

David B. Pisoni, Speech Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA. 02139, U.S.A.

Historically, the study of speech perception may be said to differ in a number of ways from the study of other aspects of auditory perception. First, the signals typically used to study the functioning of the auditory system were simple, discrete and typically differed along only a single dimension. In contrast, speech signals involve very complex spectral and temporal relations. Secondly, most of the research dealing with auditory psychophysics that has accumulated over the last thirty years has been concerned with the discriminative capacities of the sensory transducer and the functioning of the peripheral auditory mechanism. In the case of speech perception, however, the relevant mechanisms are centrally located and intimately related to more general cognitive processes that involve the encoding, storage and retrieval of information in memory. Moreover, experiments in auditory psychophysics have typically focused on experimental tasks and paradigms that involve discrimination rather than identification or recognition, processes thought to be most relevant to speech perception. Thus, it is generally believed that a good deal of what has been learned from research in auditory psychophysics and general auditory perception is only marginally relevant to the study of speech perception and to an understanding of the underlying perceptual mechanisms.

Despite these obvious differences, investigators have, nevertheless, been quite interested in the differences in perception between speech and nonspeech signals. That such differences might exist was first suggested by the report on the earliest findings of categorical discrimination of speech by Liberman et al. (1957). And it was with this general goal in mind that the first so-called "nonspeech control" experiment was carried out by Liberman et al. (1961) in order to determine the basis for the apparent distinctiveness of speech sounds.

Numerous speech-nonspeech comparisons have been carried out over the years since these early studies, including several of the contributions to the present symposium. For the most part, these experiments have revealed quite similar results. Except until quite recently, performance with nonspeech control signals failed to show the same discrimination functions that were observed with the parallel set of speech signals (Cutting and Rosner, 1974; Miller et al., 1976; Pisoni, 1977). In addition, the nonspeech signals were typically responded to by subjects at levels approximating chance performance. Such differences in perception between speech and nonspeech signals have been assumed to reflect basically different modes of perception-- a "speech mode" and an "auditory mode". Despite some attempts to explain away this dichotomy, additional evidence continues to accumulate as suggested by several of the new findings summarized in the papers included in this section.

There have been, however, a number of problems involved in drawing comparisons between speech and nonspeech signals that have raised several questions about the interpretation of the results obtained in these earlier studies. First, there is the question of whether the same psychophysical properties found in the speech stimuli were indeed preserved in the nonspeech control condition. Such a criticism seems quite appropriate for the original /do/--/to/ nonspeech control stimuli which were simply inverted spectrograms as well as the well-known "chirp" and "bleat" control stimuli of Mattingly et al. (1971) that were created by removing the formant transitions and steady-states from speech context and then presenting them in isolation to subjects for discrimination. Such manipulations while nominally preserving the speech cue obviously result in a marked change in the spectral context of the signal which no doubt affects the detection and discrimination of the original formant transitions. Such criticisms have been taken into account in the more recent experiments comparing speech and nonspeech signals as summarized by Dr. Dorman and Dr. Liberman in which the stimulus conditions remain identical across different experimental manipulations. However, several additional problems still remain in making comparisons between speech and nonspeech signals. For example, subjects in these experiments rarely if ever receive any

experience or practice with the nonspeech control signals. With complex multidimensional signals it may be quite difficult for subjects to attend to the relevant attributes of the signal that distinguish it from other signals presented in the experiment. A subject's performance with these nonspeech signals may therefore be no better than chance if he/she is not attending selectively to the same specific criterial attributes that distinguish the speech stimuli. Indeed, not knowing what to listen for may force a subject to "listen" for an irrelevant or misleading property of the signal itself. Since almost all of the nonspeech experiments conducted in the past were carried out without the use of feedback to subjects, a subject may simply focus on one aspect of the stimulus on one trial and an entirely different aspect of the stimulus on the next trial.

Setting aside some of these criticisms, the question still remains whether drawing comparisons in perception between speech and nonspeech signals will yield some meaningful insights into the perceptual mechanisms deployed in processing speech. In recent years, the use of cross-language, developmental and comparative designs in speech perception research has proven to be quite useful in this regard as a way of separating out the various roles that genetic predispositions and experiential factors play in perception. For example, while it is cited with increasing frequency that chinchillas have been shown to categorize synthetic stimuli differing in VOT in a manner quite similar to human adults, little if anything is ever mentioned about the chinchilla's failure to carry out the same task with stimuli differing in the cues to place of articulation in stops, a discrimination that even young prelinguistic infants have been shown to be capable of making. Such comparative studies are therefore useful in speech perception research to the extent that they can specify the absolute lower-limits on the sensory or psychophysical processes inherent in discriminating properties of the stimuli themselves. However, they are incapable, in principle, of providing any further information about how these signals might be "interpreted" or coded within the context of the experience and history of the organism.

Cross-language and developmental designs have also been quite useful in providing new information about the role of

early experience in perceptual development and the manner in which selective modification or tuning of the perceptual system takes place. Although the linguistic experience and background of an observer was once thought to strongly control his/her discriminative capacities in a speech perception experiment, recent findings strongly suggest that the perceptual system has a good deal of plasticity for retuning and realignment even into adulthood. The extent to which control over the productive abilities remains plastic is still a topic to be explored in future research.

To what extent is it then useful to argue for the existence of different modes of perception for speech and nonspeech signals? Some investigators such as Dr. Ades and even Dr. Massaro would like to simply explain away the distinctions drawn from earlier work on the grounds of parsimony and generality. But this is a curious position to maintain as it is commonly recognized, not only in speech perception research but in other areas of perceptual psychology, that stimuli may receive differential amounts of processing or attention by the subject, that subjects may organize the interpretation of the sensory information differently under different conditions and that the sensory trace of the initial input signal may show only a faint resemblance to its final representation resulting from encoding and storage in memory. It is hard to deny that a speech signal elicits a characteristic mode of response in a human subject-- a response that is not simply the consequence of an acoustic waveform leaving a meaningless sensory trace in the auditory periphery. Such observations suggest to me that, just as in the case of "species-typical responding" observed in the behavior of numerous other organisms, the existence of a speech mode of perception is a way of capturing certain aspects of the way human observers typically respond to speech signals that are familiar to them. Such a conceptualization does not, at least in my view, commit one to the view that human listeners cannot respond to speech in other ways more closely correlated with the sensory or psychophysical attributes of the signals themselves. To explain away the speech mode, however, is to deny the fact that a certain subset of possible acoustic signals generated by the human vocal tract are used in a distinctive and quite systematic way by both

talkers and listeners to communicate by spoken language, a species-typical behavior that is restricted, as far as I know, to homo sapiens. Past experiments comparing the perception of speech and nonspeech signals have been quite useful in characterizing how the phonological systems of natural languages have, in some sense, made use of the general properties of sensory systems in selecting out the inventory of phonetic features and their acoustic correlates. The relatively small number of distinctive features and their acoustic attributes observed across a wide variety of diverse languages suggests that the distinctions between speech and nonspeech signals still remain fundamental ones setting apart research on speech perception from the study of auditory psychophysics and the field of auditory perception more generally.

#### References

- Cutting, J.E. and Rosner, B.S. (1974) "Categories and boundaries in speech and music", Perc. Psych. 16, 564-570.
- Lieberman, A.M., Harris, K.S., Hoffman, H.S. and Griffith, B.C. (1957) "The discrimination of speech sounds within and across phoneme boundaries", J. Exp. Psych. 54, 358-368.
- Lieberman, A.M., Harris, K.S., Kinney, J.A. and Lane, H.L. (1961) "The discrimination of relative onset time of the components of certain speech and non-speech patterns", J. Exp. Psych. 61, 379-388.
- Mattingly, I.G., Liberman, A.M., Syrdal, A.K. and Halwes, T.G. (1971) "Discrimination in speech and non-speech modes", Cogn. Psych. 2, 131-157.
- Miller, J.D., J.D., Wier, C.C., Pastore, R., Kelly, W.J. and Dooling R.J. (1976) "Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception", JASA 60, 410-417.
- Pisoni, D.B. (1977) "Identification and discrimination of the relative onset of two component tones: Implications for voicing perception in stops", JASA 61, 1352-1361.

## SPEECH &amp; NON-SPEECH: WHAT HAVE WE LEARNED?

Anthony E. Ades<sup>1</sup>, Max-Planck-Gesellschaft, Projektgruppe für Psycholinguistik, Nijmegen, Netherlands.

There have been two strands of research in the speech/non-speech controversy. Firstly there are experiments where speech is compared to non-speech signals that have critical acoustic properties of speech. (See Wood, 1976, for references). This work has shown that there is no real difference: the perceptual properties of speech arise from its acoustics, not from its "speechlikeness".

This paper is concerned with the second strand, where series of speech sounds are compared to stimuli that differ along simpler dimensions like pitch and intensity. I shall summarise the arguments presented in a recent theoretical article (Ades, 1977), and extend them to other paradigms. The conclusion I shall draw is that speech/non-speech differences, as well as consonant/vowel differences, do not result from any inherent property of the sounds themselves, such as their speechlikeness, or degree of "encodedness" (Lieberman, Mattingly and Turvey, 1972), but instead depend on a property of the ensembles of stimuli used in these experiments.

This property is the range, or width of context, of the ensemble. One may think of it as the number of just-noticeable-differences across the series. This analysis is borrowed from Durlach and Braida's (1969) quantitative theory of intensity resolution. They and their colleagues, in the course of testing this theory, have obtained results for intensity that are quite analogous to results commonly obtained for speech.

There has been a consistent failure to control for the range variable when making comparisons between vowels, consonants, speech, and non-speech.

#### Identification and Discrimination

It all started, I think, with Miller's observation (1956) that we can discriminate far better than we can identify. Consider an intensity discrimination experiment where the subject is asked to decide if two sounds are the same or different. This can be done reliably if they are about one dB apart. Now suppose that there are 15 sounds, evenly spaced along a continuum spanning 25 dB. About

(1) This paper was prepared while the author held a Fellowship under the Royal Society European Science Exchange Programme.

25 discriminations could be made in this space. But when the subject is asked to label the stimuli with a number between 1 and 15 (give as much practice and feedback as possible), only seven plus or minus two categories can be used accurately. The subject is distinguishing between adjacent stimuli in identification less than half as well as (s)he would in discrimination. How can this be? After all, sensitivity to acoustic signals and to differences between them must be the same in both situations.

The answer must lie in the memory requirements. In discrimination there are two or more stimuli: the subject must store their sensory traces, compare them (perhaps by subtraction), and pronounce on the difference if any. Call this the trace mode. In the identification case, a single sensory trace must be compared to some representation of the entire stimulus series. Where, the subject must decide, does this stimulus fit, given all the others I have heard. This is the context coding mode. Presumably, the representation of the series is not in the form of traces, but is in some verbal or numerical code.

Now consider what happens to identification when the range of the ensemble is increased from 25 to 50 dB. A reasonable guess is that "accuracy", defined as the ability to place the current stimulus in context, expressed as a percentage of the size of the context, will remain constant. Of course, the absolute size of errors, expressed in j.n.d. terms, will now be larger. An archer who remains a constant 3 degrees off centre will show a larger absolute error at 50 yards than at 25.

So far, then, we assume that in discrimination (in its ideal form), the only factor affecting performance is the noisiness of the representation of the acoustic traces. In identification there will be trace noise too, but there will also be context coding noise when the subject attempts to locate a stimulus in its context. This will increase as the range increases.

The critical prediction is that as long as range is small, identification performance will be as good as discrimination. For, context noise will be minimal, and trace noise will be the only determinant in both tasks.

The theorising above is an informal statement of Durlach and Braida's (1969) quantitative theory for intensity resolution. The

above prediction, that as range decreases, identification improves, and finally approximates discrimination, was confirmed by Pynn, Braida, and Durlach (1972), for stimuli differing in intensity.

And now to speech. The classic result (Studdert-Kennedy et al., 1970) is that for series of consonant - vowel stimuli, discrimination is scarcely better than identification. Given Miller's paper on how identification is relatively weak in non speech, it was natural to see the speech results as evidence for a speech-specific mode of processing. The alternative I propose is simply that CV series are not unusual by virtue of being speech, but simply have relatively small ranges. I have shown elsewhere (Ades, 1977) that the best estimates of the range of CV series make them comparable in j.n.d. terms to the small ranges used by Durlach, Braida, and their colleagues for intensity resolution experiments.

Typically, a series of synthetic speech sounds from /ba/ to /da/, or from /ba/ to /pa/, spans between 3 and 5 j.n.d.s. A series of vowels, on the other hand must stretch across about 10 j.n.d.'s to reach from one category to another. We thus expect that discrimination on vowels will far exceed what would be predicted from identification. This has been consistently found. Generally, though, it has been interpreted to mean that vowels are somehow less "speech-like" than consonants (as if one could have stop consonants without vowels!). It should be clear that I am trying to replace this rather mystical theorising with the idea that speech, non-speech, vowels and consonants are all the same. The observed differences are due to the range variable. The number of j.n.d.'s across the series, can be used as a stimulus-free approximation of the size of the range.

#### More complex experiments

In certain cases, the range has an effect in discrimination experiments, not just in identification. This is because certain variations in the task parameters may make it profitable for the subject to operate in the context mode, i.e. to do identification: as, for instance, when the procedure adds noise or interference to the sensory trace mode. We can predict that any manipulation that makes comparison of traces harder will only worsen performance if the range is large! For, if the range is small, the subject can escape the trace noise, slip into the context-coding mode and not

suffer too much from context noise.

In these cases it is important how discrimination is tested: if the pair to be discriminated randomly changes from trial to trial, then the effective range is the range of the entire series. But if the same pair is tested many times before another part of the series is tested, the effective range will obviously be very small. It turns out that in speech research the "roving level" method is always used. Thus procedures that cause trace comparison to be harder, such as increasing the time interval between the two stimuli, or by forcing the subject to compare three traces at a time rather than two, such procedures will, in roving level testing, make the range variable critical.

Experiments of this type have been done with vowels and consonants (Pisoni, 1973, 1975). As we predict from the Durlach and Braida model, manipulations that worsen discrimination have a stronger effect on vowels than on consonants, because, according to the range hypothesis, the small range of consonants makes escape into context-coding possible without running into context memory noise. In intensity resolution, Berliner and Durlach (1973) have shown that increased time delay between stimuli to be discriminated worsens resolution only if the range is large.

#### The "anchor" effect and RT Experiments

The same ideas can be applied to other paradigms where speech and non-speech have been contrasted. In the two areas that follow I confess to being less certain of my argument, because I do not know of research where the range variable has been systematically studied.

Firstly, the "anchor effect". A series of sounds varying in pitch is constructed and the subject asked to identify them as "High" or "Low". If an endpoint stimulus (the anchor), say the highest pitched one, is presented two or three times as often as the others, the entire identification curve is shifted towards to the anchor. However, such shifts do not occur in stop-consonant series (Sawusch and Pisoni, 1973; Simon and Studdert-Kennedy, 1978). Again, we might expect that the different ranges of pitch and consonant series are involved. We may assume a 5 j.n.d. range for the speech series. The pitch series went from 114 Hz to 150 Hz: assuming a difference limen of 0.5 Hz for pitch (Klatt, 1973), this



series would span over 50 j.n.d.s. Both Simon et al., and Sawusch et al. (1974) also found a strong anchor effect in a series varying in intensity. This covered 18 dB in one experiment and 24 in the other, about 20 j.n.d.s.

Certainly, then, the range differences between the speech and non speech series were marked. But why should the range determine the anchor effect? I have no formal answer to this, but it is clear that anchor effects cannot be located in the trace mode. Also, Berliner, Durlach and Braida (1977) have shown that the "edge effect", whereby resolution in identification is better at the ends of a continuum than in the middle, and which is identified in their model as a perceptual anchoring effect in the context coding mode, is enhanced by increased range.

A second paradigm is a Reaction Time task where the subject must press one of two buttons depending on whether the stimulus is /ba/ or /da/, or whether it has high or low pitch. The point here is that if the subject is responding to the speech distinction, irrelevant variation in pitch slows the RT. However, irrelevant variation in place of articulation has a much smaller effect on RT to the pitch distinction (Day and Wood, 1972). Wood (1973) also showed that there was mutual interference between pitch and intensity, and also between place of articulation and voicing. This was interpreted as revealing two separate systems: such that there was interference within each, speech with speech, non-speech with non-speech; but no interference between.

The alternative is that both pitch and intensity discriminations are easy, while both place and voicing are harder. Interference will occur if the irrelevant variation is as salient or more salient than the distinction being tested. The situation where interference is least is precisely the one where the discrimination (pitch) is much more salient than the interfering dimension (place).

Finally, let me add that the point I have been trying to make for discriminations vs identification, anchor effects, and RT experiments has already been forcefully made for experiments on the Precategorical Acoustic Store (PAS), and on the hemispheric lateralisation of speech. The fact that sets of stop-consonant-vowel syllables produce no recency effect in PAS, whereas sets of vowels do, has been taken to mean that consonants and vowels are differen-

tially "encoded" (Liberman et al., 1972). But Darwin and Baddeley (1974) have shown that the vowel/consonant distinction here is irrelevant: what controls the recency effect is, again, the discriminability of the items within the ensemble. Similarly, the same factor is critical in determining the degree of hemispheric lateralisation for vowels (Godfrey, 1974).

#### Conclusions

At the very least it must be conceded that explorations of speech/non-speech and vowel/consonant differences might be meaningless unless factors corresponding to discriminability across the stimulus ensemble are controlled. It is obvious that the range variable is all-important in the experiments briefly reviewed here. In addition, once range is controlled for, a single unified theory for all stimuli seems well within reach. And this is surely preferable to one theory for non-speech, a second theory for consonants, (and an in-between theory for vowels).

Whether or not the above proposals are correct, the entire speech/non-speech issue seems to have acquired a life of its own, which it fights for against all odds. However, according to the views expressed here, it has taught us very little, and has simply served to direct out attention from the real problems of speech perception, exemplified for example in automatic recognition (Klatt, 1977, for a review), where the psychological contribution remains slight and engineering solutions prevail.

#### References

- Ades, A. E. (1977): "Vowels, Consonants, Speech, and Nonspeech", *Psych. Rev.* 84, 524-530.
- Berliner, J. E., and N. I. Durlach (1973): "Intensity Perception IV: Resolution in Roving Level Discrimination", *JASA*, 53, 1270-87.
- Berliner, J. E., N. I. Durlach, and L. D. Braida (1977): "Intensity Perception VII. Further Data on Roving Level Discrimination and the Resolution and Bias Edge Effects", *JASA*, 61, 1577-85.
- Darwin, C. D., and A. D. Baddeley (1974): "Acoustic Memory and the Perception of Speech", *Cogn. Psych.* 6, 41-60.
- Day, R. S., and C. C. Wood (1972): "Interaction between Linguistic and Nonlinguistic Processing", *JASA*, 51, 79(A).
- Durlach, N. I., and L. D. Braida (1969): "Intensity Perception I. Preliminary Theory of Intensity Resolution", *JASA*, 46, 372-83.



- Godfrey, J. J. (1974): "Perceptual Difficulty and the Right-Ear Advantage for Vowels". Brain and Language, 4, 323-36.
- Klatt, D. H. (1973): "Discrimination of Fundamental Frequency Contours in Synthetic Speech: Implications for Models of Pitch Perception", JASA, 53, 8-16.
- Klatt, D. H. (1977): "Review of the ARPA Speech Understanding Project", JASA, 62, 1345-66.
- Liberman, A. M., I. G. Mattingly, and M. T. Turvey (1972): "Language Codes and Memory Codes". In A. W. Melton and E. Martin (Eds.) Coding Processes in Human Memory, Washington, D. C.: Winston.
- Miller, G. A. (1956): "The Magical Number Seven, Plus or Minus Two: Some Limits on our capacity for Processing Information", Psych. Rev., 63, 81-97.
- Pisoni, D. B. (1973): "Auditory and Phonetic Codes in the Discrimination of Consonants and Vowels", Perc. Psych., 13, 253-60.
- Pisoni, D. B. (1975): "Auditory Short-Term Memory and Vowel Perception", Memory and Cognition, 3, 7-18.
- Pynn, C. T., L. D. Braida, and N. I. Durlach (1972): "Intensity Perception III. Resolution in Small-Range Identification", JASA, 51, 559-66.
- Sawusch, J. R., and D. B. Pisoni (1973): "Category Boundaries for Speech and Nonspeech Sounds", JASA, 54, 76(A).
- Sawusch, J. R., D. B. Pisoni and J. E. Cutting (1974): "Category Boundaries for Linguistic and Nonlinguistic Dimensions of the Same Stimuli", JASA, 55, S55(A).
- Simon, H. J., and M. Studdert-Kennedy (1978): "Selective Anchoring and Adaption of Phonetic and Nonphonetic Continua", JASA, 64, 1338-57.
- Studdert-Kennedy, M., A. M. Liberman, K. S. Harris, and F. S. Cooper (1970): "Motor Theory of Speech Perception: A Reply to Lane's Critical Review", Psych. Rev., 77, 234-49.
- Wood, C. C. (1973): "Levels of Processing in Speech Perception. Neurophysiological and Information Processing Analyses". Unpublished Doctoral Dissertation, Yale University.
- Wood, C. C. (1976): "Discriminability, Response Bias, and Phoneme Categories in Discrimination of Voice Onset Time", JASA, 60, 1381-89.

## SOME PSYCHOACOUSTIC FACTORS IN PHONETIC ANALYSIS

Pierre L. Divenyi, Veterans Administration Medical Center, Martinez, California, 94553.

From an ethological point of view, speech represents a complex acoustic stimulus that has the greatest survival value for man. Physically speaking, speech is complex in two ways: its spectral composition, over any epoch of arbitrary length, is extremely rich, and this spectral composition is continuously varying over time. The information density represented by the speech signal is enormous; yet, the human auditory system, despite its limited capacity, is able to receive and decode such a complex signal with remarkable efficiency. The desire to provide a reasonable explanation for such efficiency, as well as the need for descriptive data on the perceptual processes that permit reception and decoding of speech, provided much of the motivation behind the greatest part of the speech perception research accomplished to date. The emerging body of experimental findings, in turn, has constituted the background for a number of theories and models of speech perception. The leitmotiv of many of these theories, including some major contemporary ones, is that speech represents a special acoustic signal that must be handled by the auditory system in a special way (=speech mode), involving special processes and mechanisms (=phonetic feature detectors, etc.). While the special nature of speech and speech perception processes can hardly be disputed (because of their aforementioned high survival value), some recent results demonstrating speech discrimination by young infants and animals have established the need for an alternative theoretical approach-- one that would take into account, at least to some extent, some "wired-in" properties of the auditory mechanisms. The purpose of the present paper is to invoke some basic properties of the human auditory system and to reflect on the consequences of these properties for the phonetic analysis of the speech signal.

Psychophysical reality of the speech signal

Classical psychoacoustics research and classical speech perception research have progressed on traditionally separate (and not always parallel) paths. The reasons for this divorce, considered by some cynics as permanent until quite recently, were numerous, one of them being the overwhelming concern of psychoacousticians with simple acoustic signals and peripheral auditory processes. How-

ever, for the last couple of decades, the situation has gradually changed: availability of sophisticated stimulus control, the growing popularity of a systems approach to perceptual problems, and interdisciplinary orientation of an increasing number of researchers have signaled the beginnings of a (hopefully) new era. Indeed, psychoacoustics appears to be no longer afraid of spectrally and/or temporally complex sound patterns and researchers seem to address with greater freedom issues involving more central processes. Thus, it has become possible to take a fresh look upon the speech signal as a stimulus to the auditory system, and to interpret its perception in terms of a certain number of discrete psychoacoustic processes. For reasons of economy, only a few major ones will be discussed here.

Peripheral analysis and time-frequency trade. Peripheral analysis of auditory signals operates under a constraint not unlike Heisenberg's Uncertainty Principle, as defined for elementary particle physics. According to this principle, in any given system frequency resolution ( $\Delta f$ ) can be traded for temporal resolution ( $\Delta t$ ) and vice versa, such that their product  $\Delta f \Delta t$  remains constant. In the ear, such a relation is generally true only within certain limits (McGill, 1968); spectral resolution is limited by the Critical Bands (roughly 1/4 to 1/3 octaves in width; Zwicker et al., 1957) and temporal resolution by the ear's "time window" (a time constant of roughly 8 msec; Penner, 1978). Within these limits, however, this principle predicts that, to increase resolution in the spectral domain, temporal resolution must be sacrificed, and vice versa (Ronken, 1971). The validity of this prediction is proved by experimental results: discrimination of the frequency of pure tones deteriorates as their duration decreases (Moore, 1973) and, conversely, perception of the fine temporal structure of the stimulus is possible only for wide-band signals (Green, 1971).

Thus, the length of the effective time window and the width of the effective internal filter continuously adapt themselves to the spectral-temporal characteristics of the stimulus. The outcome of such an analysis will be a sequence of "neural spectra" (Klatt, 1978) or "central spectra" (de Boer, 1977) -- a series of quasi-stationary auditory events of variable duration. The temporal constraint signifies that peripheral analysis of acoustic (speech or non-speech) signals cannot be extended beyond the duration of these auditory events.

Pitch perception. According to contemporary theories (Plomp, 1975), pitch of complex signals is extracted by periodicity analysis of the internal spectrum (i.e., by taking its Fourier transform). Thus, any complex signal gives rise to two different pitch experiences: a "spectral pitch" (=formant analysis) and a "virtual pitch" (or low pitch or residue pitch [=periodicity analysis]), the former being a prerequisite for the latter. The existence region of virtual pitch is limited to pitch periods not shorter than about 2 msec ( $< 500$  Hz); the degree of its salience is a composite function of the spectral region (formant region), the serial number and the relative intensity of the component harmonics, and the periodicity rate itself (Ritsma, 1962). In complex signals consisting of several consecutive harmonics virtual pitch is determined by the eight lowest harmonics, especially those around the third (Houtgast, 1974, 264), but, interestingly, the fundamental is not dominant.

Virtual pitch is not an absolute concept: it reflects a statistical approximation to a periodicity that derives from the ensemble of peaks in the internal spectrum (de Boer, 1977). It has also been proposed (Terhardt, 1974) that virtual pitch actually represents a Gestalt property of complex sounds -- a property that is as much a result of learning as that of purely sensory processes. Such a hypothesis helps account for some systematic pitch shift phenomena that are otherwise difficult to interpret.

Temporal organization. Since peripheral analysis is limited to short temporal intervals, the sequences of "neural spectra" which temporally-complex signals generate must be organized into perceptually meaningful units by some higher-level auditory center(s). Such a perceptual organization in time obeys rules that are reminiscent of the Gestalt principles that govern the perception of visual figures in space (e.g., law of closure, law of proximity, etc; Koffka, 1935) and, ultimately, leads to the percept of an auditory pattern (Divenyi and Hirsh, 1978). Among the general rules of auditory pattern perception there is one of primary importance: two successive auditory events can be optimally resolved in time only if they occur in identical spectral bands. For example, auditory discrimination of short (10-30 msec) intervals defined by the onsets of two brief tones gradually deteriorates when the two tone frequencies become increasingly different (Divenyi and Sachs, 1978). Similarly, recognition of the temporal order of successive tones remains accurate only as long as all tone frequencies are within the same nar-

row band -- otherwise the sequence breaks into separate "auditory streams" (Bregman and Campbell, 1971).

The concept of "listening bands". The three above mentioned limitations, i.e., trade-off of time resolution - frequency resolution, limits of periodicity analysis, and restriction of accurate temporal organization to auditory events within the same narrow spectral band, are generally valid for the processing of any auditory signal, simple or complex. However, since speech constitutes an auditory stimulus in which the spectral information is generally distributed over several bands (specific to a given phonetic unit), its processing will be further complicated by yet another limitation: the auditory system is unable to simultaneously monitor several bands without loss of information (Green, 1961). The consequence of such a limitation is that auditory processing along various acoustic dimensions will be degraded by frequency uncertainty, i.e., by leaving the listener in doubt as to the frequency region in which the forthcoming auditory event is to appear. For example, frequency uncertainty will degrade detection (Creelman, 1972) and frequency discrimination (Watson, 1976) of a pure tone, as well as recognition of temporal-order patterns of several successive tones (Divenyi and Hirsh, 1978).

In order to overcome the effect of frequency uncertainty, the auditory system tends to spontaneously "tune" its focus of listening to the narrow band at or around the input frequency; it will usually remain focused at this listening band in the absence of any stimulus for at least several seconds (Johnson, 1978). Thus, at any given time, the auditory system's choice of a listening band is determined by the frequency characteristics of the last input. One of the possible reasons for the detrimental effect of frequency uncertainty is that shifting the listening focus from one band to another seems to take time (Divenyi and Hirsh, 1972). Moreover, attending to more than one spectral band at once will also degrade listening efficiency (Swets, 1963) -- the information processing capacity of the ear is, indeed, quite limited. The surprising finding is that the listener's knowledge with regard to the frequency of the forthcoming stimulus is not sufficient to completely eliminate the frequency uncertainty effect: to tune the listening band to a new region some sound (i.e., a cue) must occur (Johnson, 1978).

The locus of the tuning mechanism most probably lies above the auditory periphery: contralateral cues, too, have been found to be

effective in establishing the listening bands (Gilliom et al., 1979)

#### Relevance to speech perception

The question of great interest to many is how a system having the properties described above is likely to behave when confronted with a speech signal. While a great deal more experimental data than what we have to date are needed to answer this question (even in a marginally acceptable manner), it is nonetheless possible to give a cursory outline of the effects of auditory processing on speech sounds. Again, because of space limitations, the picture presented here will be sketchy and less than exhaustive.

Segmentation. As a direct result of the time resolution - frequency resolution trade-off, any complex signal in which narrow-band and wide-band portions alternate will be automatically segmented at a peripheral level. Since, in speech, transitions from wide-band to narrow-band acoustic segments (and vice versa) roughly correspond to phonetic segment dividers, each of these transitions (smoothed by the ear's time window function) will produce marker signals at the auditory periphery. Thus, the series of auditory events (= "neural spectra") which some higher-level centers will organize into perceptual units will actually be a succession of phonetically meaningful elements.

Speaker invariance. The mutual interdependence of waveform periodicity, spectrum of complex sounds, salience of virtual pitch, and salience of spectral pitch can account for much of the formant frequency - fundamental frequency relations observed in vowel production and perception (Fujisaki and Kawashima, 1968). Since vowels (=quasi-steady-state sounds) are analyzed in a narrow-band mode, relatively small spectral variations may be detected by the auditory system. Such a large degree of sensitivity may provide the explanation underlying the notion that vowel perception is "continuous" rather than "categorical".

Categorical perception and selective adaptation. In CV syllables, especially in stop-vowel pairs, the initial consonant is a wide-band transient; therefore, nothing compels the auditory system to tune the listening band to any particular position of the spectrum. The relative freedom of tuning that derives from wide-band stimuli enables the auditory system to select a frequency region to which it will spontaneously direct its focus before the onset of the CV sound. Such strategies may possibly originate in learning:

category boundaries that characterize certain features are known to be language-bound. However, strategies for positioning the listening band are by no means absolute: a sound of different spectral-temporal characteristics (speech or non-speech, see Samuel and Newport, 1979) presented prior to the CV stimulus could serve as a cue (Johnson, 1978) and make the auditory system choose a different listening band. Thus, selective adaptation effects could be re-interpreted in terms of pre-cueing and listening bands.

Such an interpretation is quite straightforward when one looks upon category boundary shifts observed for the feature of place-of-articulation in adaptation experiments: the acoustic basis for this feature is almost exclusively spectral. Explanation of boundary shifts of the voiced-voiceless category, a predominantly temporal feature, is somewhat more complex. Since temporal organization of acoustic events heavily depends on temporal cues contained in some narrow band, perception of the feature of voicing will be a function of the discriminability of voice-onset-time inside one (or several) narrow spectral region(s). However, when a brief auditory time interval is marked by a pair of sounds of identical spectral composition, temporal masking (forward or backward) of one marker by the other could decrease the discriminability of the interval (Divenyi and Sachs, 1978). Because the relative energy of the consonant and the vowel varies from one band to another (thereby also causing the amount of temporal masking to vary), the choice of the monitored band will be critical in determining the VOT boundary. Thus, tuning the listening band to different spectral regions will result in different voicing boundaries. An adaptor stimulus (by virtue of its potential role as a cue), therefore, may alter the natural position of the listening band for a given CV syllable, thereby producing a shift in the category boundary. It is conceivable that perceptual-productive acquisition of different phonetic patterns could also be associated with different spectral positions that the listening band will spontaneously occupy; thus, the present theory is consistent with the language-dependent nature of voicing category boundaries.

Time invariance implies that the relative duration of certain phonetic segments is irrelevant. Experiments on the perception of non-speech sound sequences (Watson, 1976; Divenyi and Hirsh, 1978) have shown that the emergence of an auditory pattern (at least within certain limits) does not depend on the absolute duration of the

components. Thus, it follows that the rate at which the speech segments ("neural spectra") of the speech sounds occur will not change the "figural properties" of the patterns.

#### Conclusion: Whither phonetic analysis?

When attempting to examine speech processing on the auditory level, one finds that the product of auditory analysis possesses several characteristics that are customarily thought to belong to the realm of phonetic analysis (feature analysis, etc.). While it is readily acknowledged here that many crucial experiments needed to prove (or disprove) critical points have not yet been performed, and that straight extrapolation of non-speech auditory data to speech-bound processes may often be risky, we feel, nevertheless, that auditory analysis of the speech signal well exceeds the limits imposed on it by several widely accepted theories. The view that phonetic analysis may not be an indispensable stage in speech processing is concordant with the opinion expressed in some studies on the perception of speech by man (Ades, 1976) or the recognition of speech by machine (Klatt, 1978). An alternative view, one that we would like to propose herewith, is that speech perception may be regarded as a special class of auditory pattern perception -- special only because we have learned these patterns so well.

#### References

- Ades, A.E. (1976): "Adapting the property detectors for speech perception", in New approaches to language mechanisms, R.J. Wales and E. Walker (eds.), 55-108, Amsterdam: North Holland.
- de Boer, E. (1977): "Pitch theories unified", in Psychophysics and physiology of hearing, E.F. Evans and J.P. Wilson (eds), 323-334, London: Academic.
- Bregman, A.S. and J.L. Campbell (1971): "Primary auditory stream segregation and perception of order in rapid sequences of tones", JEP 89, 244-249.
- Creelman, C.D. (1972): "Detecting signals of uncertain frequency: Analysis by individual alternative signals", JASA 52, 167.
- Divenyi, P.L. and I.J. Hirsh (1972): "Discrimination of the silent gap in two-tone sequences of different frequencies", JASA 51, 138.
- Divenyi, P.L. and I.J. Hirsh (1978): "Some figural properties of auditory patterns", JASA 64, 1369-1385.
- Divenyi, P.L. and R.M. Sachs (1978): "Discrimination of time intervals bounded by tone bursts", Perc. Psych. 24, 429-436.
- Fujisaki, H. and T. Kawashima (1968): "The roles of pitch and higher formants in the perception of vowels", IEEE AEA AU-16, 73-77.
- Gilliom, J., D.W. Taylor and C. Cline (1979): "Timing constraints

- for effective cueing in the detection of sinusoids of uncertain frequency", Perc. Psych. 25 (in press).
- Green, D.M. (1961): "Detection of auditory sinusoids of uncertain frequency", JASA 33, 897-903.
- Green, D.M. (1971): "Temporal auditory acuity", Psych. Rev. 78, 540-551.
- Houtgast, T. (1974): "Masking patterns and lateral inhibition", in Facts and models in hearing, E. Zwicker and E. Terhardt (eds.), 258-265, Berlin: Springer.
- Johnson, D.M. (1978): "Attentional factors in the detection of uncertain auditory signals", Unpubl. Doct. Dissert. Univ. Calif. Berkeley.
- Klatt, D.H. (1978): "Speech perception: A model of acoustic-phonetic analysis and lexical access", in Perception and production of fluent speech, R.A. Cole (ed), Hillsdale (N.J.): Erlbaum.
- Koffka, K. (1935): Principles of Gestalt psychology, New York: Harcourt Brace.
- McGill, W.J. (1968): "Polynomial psychometric functions in audition", J. Math. Psych. 5, 369-376.
- Moore, B.C.J. (1973): "Frequency difference limens for short-duration tones", JASA 54, 610-619.
- Penner, M.J. (1978): "A power-law transformation resulting in a class of short-term integrators that produce time-intensity trades for noise bursts", JASA 63, 195-201.
- Plomp, R. (1975): "Auditory psychophysics", Ann. Rev. Psych. 26, 207-232.
- Ritsma, R.J. (1962): "Existence region of the tonal residue", JASA 34, 1224-1229.
- Ronken, D.A. (1971): "Some effects of bandwidth-duration constraints on frequency discrimination", JASA 49, 1232-1242.
- Samuel, A.G. and E.L. Newport (1979): "Adaptation of speech by non-speech: Evidence for complex acoustic cue detectors", JEP HPP 5 (in press).
- Swets, J.A. (1963): "Central factors in auditory frequency selectivity", Psych. Bull. 60, 429-440.
- Terhardt, E. (1974): "Pitch, consonance, and harmony", JASA 55, 1061-1069.
- Watson, C.S. (1976): "Factors in the discrimination of word-length auditory patterns", in Hearing and Davis: Essays honoring Hal-lowell Davis; S.K. Hirsh, D.E. Eldredge, I.J. Hirsh, and S.R. Siverman (eds.), 175-189, St. Louis: Washington University Press.
- Zwicker, E., G. Flottorp and S. S. Stevens (1957). "Critical band width in loudness summation", JASA 29, 548-557.

ON THE IDENTIFICATION OF SINE-WAVE ANALOGUES OF CV SYLLABLES<sup>1</sup>

Michael F. Dorman,<sup>2</sup> Haskins Laboratories, 270 Crown Street,  
New Haven, Connecticut 06510, United States of America

In order to answer the question - Do infants perceive speech phonetically? - the stimulus continuum presented to the subjects must have phonetic category boundaries which are clearly dissociated from auditory category boundaries. For, if the two boundaries coincide, then the subjects' basis for response can not be determined. This situation, in the view of several authors, characterizes the identification of categories along the voice-onset-time (VOT) continuum. For example, Pisoni (1977) suggests that the auditory categories of simultaneous and nonsimultaneous onset could underlie infants' discrimination along the VOT continuum.

In the present series of experiments our aim was to determine whether auditory categories may also underlie infants' ability to discriminate between stop consonants which differ in place of articulation. An examination of the stimuli used in Eimas' (1974) and Miller and Morse's (1976) studies of infant place discrimination suggests a possible psychoacoustic basis for the discrimination between [bae] and [dae] - i.e., the discrimination could be based on the difference between frequency change and no frequency change in the second and third formant transitions. While the outcomes of the two studies lend little support to this position, we felt, nevertheless, that it would be important for future research to assess whether auditory categories generally coincide with phonetic categories along a continuum of  $F_2$  and  $F_3$  change.

The procedure used in our experiments was to present adults with consonant-vowel (CV) syllables synthesized with formant structure and CV analogues synthesized with frequency and amplitude modulated sine waves. Our rationale for this approach was that if listeners placed category boundaries at the same place along both the speech and nonspeech continua, then we should believe that, for these stimuli at least, the phonetic category boundaries coincide with acoustic category boundaries. These stimuli, of course, would be inappropriate for use with infants. If, on the other hand,

- 
- (1) This research was a collaborative effort among Dr. Peter Bailey, Dr. Quentin Summerfield and myself.
  - (2) Also, Arizona State University, Tempe, Arizona 85281, United States of America.



the two boundaries did not coincide, then the speech stimuli could well prove probative in studies of infant speech perception.

The stimuli for our first experiment were a [bo-do] continuum and a [be-de] continuum (see Figure 1). The first and third formants in both continua were identical - only the second formant differed between the two. The parameter values were selected so that both continua would be physically symmetrical but phonetically asymmetrical. We intended the phoneme boundary along the [bo-do] continuum to be associated with a falling transition so that the majority of the stimuli would be heard as [bo]. In contrast, we intended the phoneme boundary along the [be-de] continuum to be associated with rising transitions so that the majority of the stimuli would be heard as [de]. In this way we intended to dissociate phonetic boundaries from auditory boundaries that may accompany flat as opposed to rising or falling transitions, or from auditory boundaries that might simply coincide with the center of the stimulus range.

To generate identification functions for these stimuli, we presented the stimuli to our listeners in an AXB format. On each trial three stimuli were presented; the first and third members of the triad were the end points of the continuum, the second member was a stimulus drawn randomly from that continuum. The task of the listeners was to indicate whether the second stimulus was more like the first or more like the last member of the triad. We chose this task to avoid a problem usually associated with the absolute identification of nonspeech stimuli - that listeners have more difficulty attaching category labels to the nonspeech stimuli than to the speech stimuli. By presenting the end points of the stimulus continuum on each trial in both the speech and nonspeech conditions, we hoped to make the identification task equally difficult in both conditions.

Turning now to the result of our first experiment, we see in Figure 2 the identification function for the speech signals. As predicted there were more [b]-like responses for the [bo-do] continuum than for the [be-de] continuum. However, the difference between the locations of the phoneme boundaries fell short of significance. In contrast to the largely asymmetric identification functions shown for the formant stimuli, the identification functions for the sine-wave analogues, shown in Figure 3, coincide

throughout their range. We would conclude from this outcome that at least the [bo-do] boundary does not coincide with an acoustic category boundary. There are, however, two possible interpretations of this outcome: the asymmetrical categorization of the formant continua could either be correlated with the way the stimuli are heard - as speech or nonspeech - or may simply be correlated with the different spectral properties of the formant and sine-wave stimuli. To rule out the latter possibility we would like the same physical signal to be heard as speech-like in one context and as nonspeech in another. If the category boundaries differed in this instance then it certainly could not be argued that spectral differences account for the outcome. Fortunately, the sine-wave stimuli used in our experiment fit this requirement nicely. After we instructed our listeners as to the nature of the sine-wave stimuli, they readily agreed that the stimuli could be heard as stop initiated.

The outcome of this experiment (when the sine waves were heard as speech) is shown in Figure 4. The pattern of results is clearly very different from that when the sine waves were heard as nonspeech. Here, the two functions no longer overlap. As with the formant stimuli, the majority of the [bo-do] analogues were heard as [b]-like. Moreover, the category boundaries along the two continua differed significantly. It is clear that the pattern of results obtained when the sine waves were heard as speech-like is more akin to that obtained for the formant stimuli than for that obtained when the sine waves were heard as nonspeech. It appears, then, that the difference between the speech and nonspeech conditions was not due to the spectral differences as such, but, rather, was due to the way in which the stimuli were heard.

To assess the reliability of our first experiment we conducted a second experiment. For this experiment we synthesized a single [ba-da] continuum and a corresponding nonspeech analogue with sine waves. The speech continuum was more natural sounding than either of those used in our first experiment and, perhaps as a consequence, many listeners heard the sine-wave analogues as speech-like without prompting. Thus, we were able to divide our subjects into two groups on the basis of their perception of the sine-wave stimuli.

The identification function for the subjects who heard the sine waves as speech is shown in Figure 5 along with the identifica-



tion function for the formant stimuli. The two functions are quite similar and two phoneme boundaries fall to the right of the mid-point of the stimulus continuum. In contrast, the identification function for the subjects who heard the sine waves as nonspeech appears quite different from that generated in response to the formant stimuli (see Figure 6). This difference is reflected in the significantly flatter slope of the nonspeech function and fewer [b]-like responses to the nonspeech stimuli.

#### Summary

Earlier in this paper we raised the question of whether auditory category boundaries generally coincide with phonetic category boundaries along  $F_2$ - $F_3$  continua. Unfortunately, our results provide an equivocal answer; the [bo-do] boundary in our first experiment did not seem to coincide with the auditory boundary, but the phonetic and auditory boundaries in the second experiment were uncomfortably close. Nevertheless, we have gained a significant purchase on a methodology that will allow us to dissociate auditory and phonetic boundaries. We see, then, an opportunity to construct continua which will be of use in the study of the ontogeny of phonetic perception.

Moreover, we see quite clearly that the perceptual system categorizes sine-wave stimuli as a function of how they are heard: when heard as speech they are categorized like formant stimuli; when heard as nonspeech they are categorized differently. We should wonder then what mechanisms underlie this changing percept of an unchanging stimulus. The nature of those mechanisms will be, I believe, the topic of Dr. Bailey's and Dr. Summerfield's paper.

#### References

- Eimas, P.D. (1974): "Auditory and linguistic processing of cues for place of articulation by infants", *Perc.Psych.* 16, 513-521.
- Miller, C.L. and P.A. Morse (1976): "The 'Heart' of categorical speech discrimination in young infants", *JSHR* 19, 578-589.
- Pisoni, D.B. (1977): "Identification and discrimination of the relative onset times of two component tones: Implications for voicing perception in stops", *JASA* 61, 1352-1361.

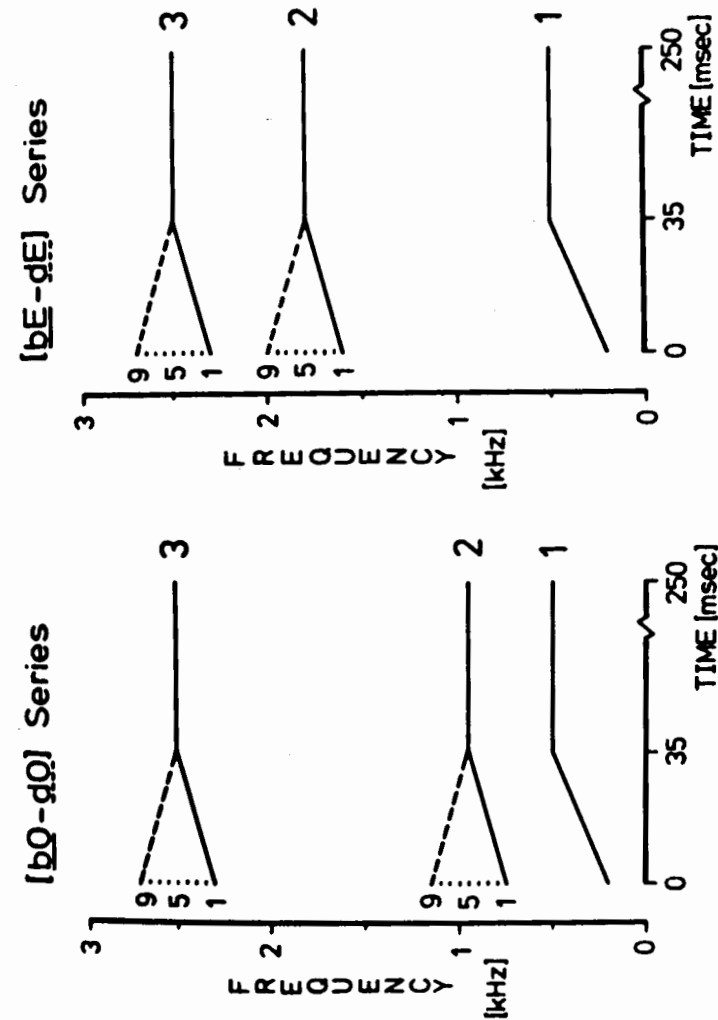


Figure 1  
Stimuli for first experiment

FORMANTS heard as SPEECH

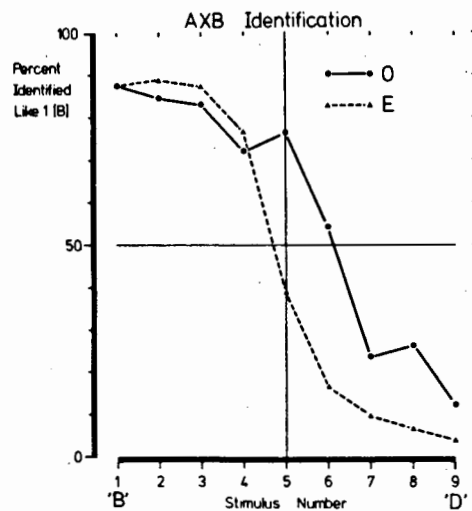


Figure 2

SINE-WAVES heard as NON-SPEECH

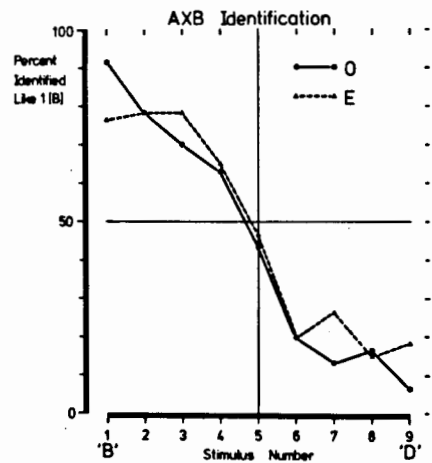


Figure 3

SINE-WAVES heard as SPEECH

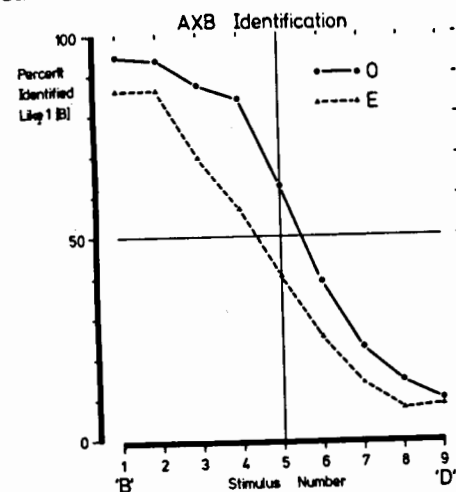


Figure 4

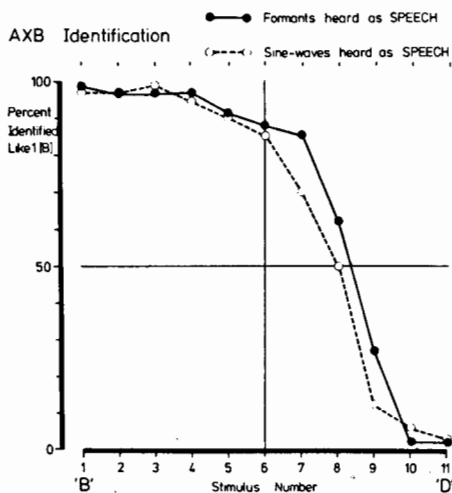


Figure 5

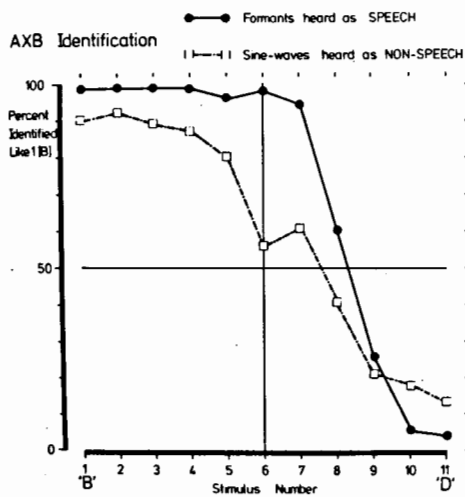


Figure 6

PERCEPTUAL LEARNING OF MIRROR-IMAGE ACOUSTIC PATTERNS<sup>1</sup>

M.E. Grunke and D.B. Pisoni<sup>2</sup>, Psychology Department, Indiana University, Bloomington, Indiana 47401 U.S.A.

The phonetic units of spoken language are often mapped in a one-to-many fashion onto their acoustic representations in speech. As a consequence, a child acquiring language must in some way be able to recognize a variety of different acoustic signals as members of the same phonetic category. For example, the /b/ in the syllable /ba/ is characterized by rapidly rising formant transitions into the following vowel, whereas the same consonant in the syllable /ab/ is characterized by rapidly falling formant transitions.

In view of the lack of one-to-one correspondence between phonemes and their acoustical representations in speech, it would seem advantageous for a child learning language if the various acoustical forms of a particular phoneme were related to each other perceptually. In fact, the acoustical representations of stop consonants in initial and final position, although physically different, are related: stop consonants in final syllable position are roughly the mirror image in time of their counterparts in initial position. But are mirror-image acoustic patterns inherently related perceptually for the listener?

The issue of the perceptual relatedness of mirror image acoustic patterns was addressed recently in a series of experiments by Klatt & Shattuck (1975) & Shattuck and Klatt (1976). They presented brief pure-tone acoustic patterns to adult listeners who had to make a similarity judgment. The acoustic patterns were two-component frequency glissandos with a short-term spectral composition similar to the formant transitions in speech.

The results of these experiments did not support the original hypothesis that mirror-image acoustic patterns are intrinsically similar for a listener. Instead, judgements of perceptual similarity for these patterns were based primarily on

---

1) This research was supported by NIMH research grant MH-24027-04, NIMH Post-doctoral fellowship MH-5823-03 to MEG and a fellowship from the Guggenheim Foundation to DBP.

2) Currently at the Speech Group, Research Laboratory of Electronics, M.I.T. Cambridge, Mass.

the direction of the lower glissando component, the component occurring in the region of the second formant.

We have conducted three experiments that also address the question of whether mirror-image acoustic patterns are intrinsically related. However, we used acoustic patterns that included a steady-state constant frequency (CF) portion as well as a rapid frequency glissando (FM). In addition, we used a perceptual learning paradigm in which listeners had to learn to map four different acoustic patterns into two response categories. We wanted to know which of several mapping arrangements of these patterns would be easiest for listeners to learn.









### Stimuli

Three sets of stimuli, with four signals per set, were generated using a complex-tone generating program (Kewley-Port, 1976). Each stimulus component consisted of a 60 msec linear rise or fall (FM) in frequency and a 140 msec constant-frequency (CF) portion. The four signals within a set differed in whether the frequency transition was rising or falling and whether the transition preceded or followed the steady-state portion.

The three stimulus sets differed in the number of component tones, either one, two or three. Frequency values were selected to correspond to values of the first, second, and third formants in the syllables /ba/, /da/, /ab/, and /ad/. For the Single-Tone set, the patterns corresponded to the frequency of the second formant. For the Double-Tone set, component frequencies corresponded to the second and third formants. For the Triple-Tone set, all three formants were represented although the frequency transitions corresponding to the first formant always rose when it preceded the steady-state and fell when it followed the steady-state, in accordance with the formant motions observed in natural speech.

### Experimental Procedure and Design

In the perceptual learning task one stimulus from a particular set was presented via headphones on each trial to subjects who responded by pressing one of two response buttons. Correct feedback was provided after each response according to one of three stimulus mapping arrangements. In the Mirror-Image

condition, stimuli with a rising transition preceding the steady-state (  ) or a falling transition following the steady-state (  ) were assigned to one response (R1) whereas stimuli with a falling transition preceding the steady-state (  ) or a rising transition following the steady-state (  ) were assigned to the other response (R2). In the Rise-Fall condition, stimuli with rising transitions either preceding or following the steady-state (  ) were assigned to one response: the two stimuli with falling transitions (  ) were assigned to the other. In the Temporal-Position condition, the stimuli were assigned to responses according to the temporal position of the transitions -- whether the transition preceded (  ) or followed (  ) the steady-state frequency.

In addition to test trials on which responses were collected, study periods were also interspersed to help subjects learn the appropriate stimulus-response mapping. During study periods, several repetitions of each of the test stimuli were presented while feedback was provided.

### Results and Discussion

Responses were analyzed in terms of number correct by stimulus. Of the three mapping conditions, the Temporal-Position condition showed the highest performance with 88.4% correct. More importantly, however, the Mirror-Image condition produced more accurate responding, 78% correct, than the Rise-Fall mapping condition with 68.3% correct response.





With respect to stimulus set, the Double-Tone stimuli showed slightly better performance, 81.6%, than the Single-Tone set, with 78.7%. However, the Triple-Tone set which more closely resembled speech showed the poorest performance with only 74.5% correct.

These results have several implications for learning of acoustical patterns resembling speech. First, the easiest stimulus mapping condition to learn was the one based on the temporal position of the transition in the pattern -- a relationship that clearly does not require the subject to analyze out the component frequencies at onset or offset. Subjects can simply use temporal position (i.e., initial vs. final) as the

most salient dimension for learning and ignore all other differences. Second, it is also apparent from our results that when the position of the transition becomes an irrelevant attribute to be ignored by the subject, a stimulus arrangement based on a mirror-image relationship is, in fact, easier to learn than one based on only the direction of frequency change of the transitions. Thus, while mirror-image patterns are not the same perceptually, subjects are nevertheless able to recognize and selectively attend to the criterial properties of the stimuli that define their equivalence. In both of these mapping conditions, subjects must "hear-out" the individual components of the patterns and respond to them selectively.

In Experiment 1 we used tonal patterns so that the stimuli would not be heard as speech. The results would be uninteresting if subjects simply heard these patterns as speech and used phonetic labels to mediate acquisition. By this interpretation, mirror-image patterns would be superior to direction-of-transition mapping because the stimuli within a mirror-image pair evoke the same phonetic label -- i.e. "b" or "d" -- and not because their configural properties are intrinsically related. To study the effects of categorization on perceptual learning we carried out another experiment to determine how well subjects could identify these patterns when explicitly provided with labels that emphasized attention to either the acoustic or phonetic properties of the stimuli.

#### Experimental Procedure and Design: Experiment 2

In Experiment 2 subjects were required to identify the stimuli into one of four categories provided by the experimenter. In the Acoustic-Label condition, subjects were told that the stimuli were tones consisting of a short interval with constant pitch, followed or preceded by a rapid rise or fall in pitch. The acoustic labels were schematic line drawings of the time course of the frequency change of each stimulus ( , , ,  ). In the Phonetic-Label condition, subjects were told that the stimuli were modified tokens of natural speech and were provided with the labels /ba/, /da/, /ab/, and /ad/.

Results: Experiment 2

For the Single- and Double-Tone stimuli, acoustic labels were matched more accurately than phonetic labels. This effect was reversed, however, for the Triple-Tone stimuli, where phonetic labels were more accurate than acoustic labels. The addition of a component in the region of the first formant markedly increased the accuracy of phonetic categorization while decreasing performance with acoustic labels. Note, however, that the low tone for each Triple-Tone stimulus had either an initial rising or final falling transition which did not parallel the direction of the other transitions. The presence of these conflicting frequency glissandos no doubt produced interference in the acoustic labeling condition.

In Experiment 1 acquisition performance was poorer for Triple-Tone stimuli than for Single- and Double-Tone stimuli, a pattern which was replicated here only in the Acoustic Label conditions. Thus, these results support the interpretation that subjects in Experiment 1 were listening primarily in an auditory, rather than phonetic mode and strongly suggest that phonetic-mediation was not responsible for the outcome observed earlier.

The mirror image patterns used here always shared three properties in common: (1) they had the same steady-state frequencies, (2) the frequency transitions at onset and offset had the same short-term spectral composition, and (3) members of a given mirror-image pair had roughly the same average frequency. These three properties are potential factors that could contribute to the salience of mirror-image pairs and the advantage observed in learning. The first factor could not have played a role in the earlier results since all pairs within a stimulus set, whether mirror-image or not, had identical steady-states. However, the other two properties could have been used as reliable discriminative cues by subjects to facilitate learning.

To determine which of these acoustic attributes, if any, was responsible for the advantage observed in learning mirror-image pairs, we repeated Experiment 2 adding two additional stimulus sets, one with average frequencies and the other with the



transitions adjusted to be equal. We also included the original constant steady-state stimuli. Would the mirror-image condition continue to show an advantage in learning under these two new stimulus conditions?

#### Experimental Procedure and Design: Experiment 3

The experimental procedure was the same as Experiment 1. However, only Double-Tone stimuli were used and the Temporal-Position mapping condition was eliminated. Three stimulus sets were constructed, again with four stimuli per set. Each set contained the initial-rising and final-falling stimuli used earlier. For the stimuli with identical transitions, frequency values for the initial-falling and final-rising stimuli were set lower than in the previous experiment, so that the transitions were identical for all stimuli. These two stimuli were also lowered in frequency for the Average-Frequency set to make all stimuli equal in average frequency.

#### Results: Experiment 3

The difference in performance between Mirror-Image and Rise-Fall mapping conditions was replicated with the Constant-Steady-State stimuli. The difference between Mirror-Image and Rise-Fall was, however, even greater for the stimuli containing identical transitions. However, for the Average-Frequency stimuli, no significant difference was found between the two mapping conditions. Thus, on the one hand, the advantage of Mirror-Image over Rise-Fall mapping can be made more pronounced by adjusting the frequency transitions to be identical whereas the difference can be attenuated substantially by adjusting all stimuli to have the same average frequency irrespective of the temporal and spectral properties of the patterns.

#### Conclusions

Mirror-Image acoustic patterns show an advantage in perceptual learning because subjects respond not only to individual components of these patterns but also to properties of the entire pattern in terms of its configural shape. Subjects do not seem to attend selectively to only the gross shape of the spectrum at onset or offset but prefer instead to integrate and deploy salient cues contained in both the transitional and

steady-state portion of the entire patterns. In the case of Mirror-Image patterns, criterial differences between the responses happen also to be correlated with salient and well-defined redundant properties of the patterns such as average pitch which was an irrelevant and uncorrelated dimension when the patterns were arranged in the Rise-Fall mapping condition.

It is apparent from these results with nonspeech signals having properties similar to speech that differences in "mode of processing" can also control perceptual selectivity and influence the perception of individual components of the stimulus pattern as well as the entire pattern itself. This can occur in quite different ways depending on whether the subject's attention is directed to coding the auditory properties or the phonetic qualities of the patterns. These new results on mirror-image patterns have been obtained in a perceptual learning task despite the report that the perceptual similarity of these acoustic patterns cannot be recognized consciously by subjects as shown by earlier experiments.

#### References

- Klatt, D.H. and S.R. Shattuck (1975): "Perception of brief stimuli that resemble rapid formant transitions," In G. Fant and M.A.A. Tatham (eds.): Auditory Analysis and Perception of Speech. New York: Academic Press, 294-301.
- Kewley-Port, D. (1976): "A complex-tone generating program," Research on Speech Perception Progress Report No. 3 Department of Psychology, Indiana University, Bloomington, Indiana.
- Shattuck, S.R. and D. Klatt (1976): "The perceptual similarity of mirror-image acoustic patterns in speech," Perception and Psychophysics, 20, 470-474.

DUPLEX PERCEPTION AND INTEGRATION OF CUES: EVIDENCE THAT SPEECH IS DIFFERENT FROM NONSPEECH AND SIMILAR TO LANGUAGE

Alvin M. Liberman, Haskins Laboratories, New Haven, Connecticut

The observations relevant to a comparison of speech and nonspeech -- the subject of our symposium -- can be divided, according to one's purposes, into several classes. As my contribution to our discussion, I would call your attention to two. In the one class are those research findings that can enlighten us about speech as a putative mode of perception, different phenomenologically from other aspects of audition. In the other class are those findings that permit us to see one of the possibly unique characteristics of speech as an instance of a correspondingly unique characteristic of language.

Phonetic perception as a mode. Perhaps the most direct way to observe the characteristic differences between perception in the phonetic and auditory modes is to contrive to have the same stimulus patterns perceived, either alternatively or simultaneously, as speech and nonspeech. The first reported example was by Lane and Schneider (1963). Using more or less unnatural synthetic speech that sampled the continuum of voice-onset-time from voiced to voiceless syllable-initial stops, they undertook to obtain discrimination functions under two conditions: in one, they told the subjects they would hear speech; in the other, arbitrary sounds of a complex sort. In the event, the discrimination functions were different: those obtained in the "speech" condition showed the usual peak at the phonetic boundary, while those from the other condition did not.

A more recent and thoroughgoing experiment of this general type has been carried out by Bailey, et al. (1977). Inasmuch as it is to be reported at this symposium, I will say no more about it.

In both of these studies, stimulus patterns were heard as speech or nonspeech, but not at the same time. Let us turn now to those cases in which the two percepts are experienced simultaneously. Such duplex perception was accomplished first by Rand (1974), and applied by him to perception of the transition cue for syllable-initial stops. Into one ear he put just the brief second- and third-formant transitions that are sufficient,

in the appropriate context, to produce the perceived difference in place among [b], [d], and [g]. By themselves, these transitions sound more like a nonspeech 'chirp' than anything else. Into the other ear, he put the remainder of the pattern. Let us call it the 'base'. By itself, the base sounds more or less like a syllable. To some listeners, indeed, it appears to be a stop-vowel syllable but, having a fixed acoustic structure, it can, of course, produce only one stop, not all three. The interesting effect occurs when the transition cue and the base are presented dichotically (and in approximately the right time relationship), for then the listener fuses the two inputs so as to perceive the three coherent stop-vowel syllables he would have perceived had the two inputs been mixed electronically (and presented binaurally), while at the same time perceiving the 'chirp' he would have perceived had the transition cues been presented in isolation. (The chirp is heard in the ear to which the isolated transitions are presented; the syllable is heard in the other ear.) Thus, the same brain perceives the same stimulus input in two phenomenologically different ways, as speech and nonspeech, at the same time. This provides, at the very least, an excellent way to gain an impression of what the difference between phonetic and auditory modes sounds like.

Being interested in the possibility of using Rand's technique for the purpose of further and more nearly precise comparisons of the phonetic and auditory modes, I succeeded in reproducing the phenomenon, but experienced difficulty in getting it to be sufficiently stable to permit the further investigations I had in mind. Recently, however, David Isenberg and I (1978) have produced a duplex percept like that of Rand, but easier to hear, we think, and also, perhaps, more stable. We followed Rand's procedure, but changed the stimulus pattern. In particular, we thought it advisable to make the critical (and isolated) cue be the third-formant transition. The advantage, in that case, might be that the remainder of the pattern -- all of the first and second formants, plus the steady-state portion of the third formant -- would be quite full and speechlike. Accordingly, we chose the contrast [r] vs [l], putting the critical third-formant transition cue into one ear and the remainder (the base) into the other. It was quickly apparent that this arrangement

did make it relatively easy to obtain the duplex percept. Indeed, tests with a number of listeners confirmed that they could "correctly" hear [r] or [l] (depending on which transition cue was presented), while simultaneously hearing the transition cue as a chirp.

Having thus found that the simultaneous fusion and separation of the two parts of the pattern (base and isolated transition) seemed to occur easily and consistently, we undertook further tests. The one I would briefly describe here was designed to assess the effects on the duplex percept of separately varying the intensity of the two stimulus components. We observed, first, that changing the intensity of the transition cue caused changes in the perceived loudness of the chirp, but no such changes in the fused [ra] or [la]; changes in the perceived loudness of the fused [ra] and [la] seemed to occur only as a result of variations in the intensity of the base.

To test this manifestation of duplexity more systematically, we carried out the following experiment. On each trial, we presented to the listener a sequence of dichotic pairs in which the transition component (in one ear) varied between the [r] cue and the [l] cue according to some predetermined order. Also, on each trial we varied the intensity of (1) the transition cue, or (2) the base, or (3) neither. The listener's task was to tell the order of [ra] and [la] syllables he heard, and, in addition, which loudnesses, if any, had changed. The results indicated that our listeners were, in fact, fusing the dichotic inputs to hear the 'correct' syllable, while at the same time dissociating the loudnesses by assigning the intensity of the transition cue to the chirp and the intensity of the base to the fused speech percept. We find this phenomenon interesting in its own right, but also as a basis for further investigation into the properties of the two components -- speech and nonspeech -- of the duplex percept. Consider, in that connection, an earlier study by Mattingly et al (1971) that compared the discrimination of a transition cue when, in the one case, it was presented in isolation and perceived as a chirp, and when, in the other, it was in its proper place in the speech pattern and cued a phonetic distinction. That study found differences in the discrimination pattern under the two conditions, but the interpretation of that finding

was subject to the reservation that the transition cue was, after all, in different contexts. By taking advantage of the duplex percept, we can, perhaps, obtain results that will avoid the need for that reservation and thus speak more straightforwardly to the difference between speech and nonspeech.

The integration of cues in speech and syntax. One of the most general characteristics of speech is that the information appropriate to a phonetic segment is typically contained in a numerous variety of cues; moreover, these are widely distributed through the signal and sometimes overlapped with cues for other phones. This is so because of the nature of articulation and coarticulation (Cooper, 1963; Fant, 1973): the various components of an articulatory gesture, distributed as they are in time, spread their acoustic consequences through the signal. Thus, the closing and opening gestures appropriate to an intervocalic stop affect the duration of the preceding syllable and also its offset, the occurrence and duration of an intervocalic silence, and the temporal and spectral characteristics of the onset of the following syllable. Conversely, and as a result of coarticulation, information about successive segments is often collapsed into a single acoustic segment and conveyed simultaneously, as in the case of most consonant-vowel syllables. Yet the speech processor somehow sorts the cues, as it were, assigning each to the appropriate part of the perceived phonetic structure. More to the point of our present purpose, it "integrates" into a unitary percept all the cues for a particular phone, no matter how various and widely distributed the cues may be (Lieberman and Studdert-Kennedy, 1977; Repp et al, 1978; Bailey and Summerfield, 1978; and Dorman et al, 1978). It is difficult to see how this can be accomplished by ordinary auditory mechanisms, so we assume phonetic processes specialized for the purpose.

Consider, for example, the above-mentioned experiment by Repp et al. It dealt with perception of the utterance: "Did you see the gray (great) ship (chip)?" The variables of interest were (1) the nature of the next-to-last word, which was biased either toward gray or great; (2) the duration of the silent interval between gray (great) and ship (chip), and (3) the duration of the fricative noise in ship (chip).

Let us now look first at the "forward" action of an earlier-

occurring cue on a later-occurring one: given a perceptual boundary between ship and chip that varied according to the duration of the fricative noise and also the duration of the preceding silent interval, there was a further variation that depended, other things equal, on whether the preceding word was biased toward gray or great. Now consider an effect in the opposite direction -- the "backward" action of a later-occurring cue on the perception of an earlier-occurring one. This was exemplified by the finding that the listener perceived gray or great depending, all else equal, on the duration of the fricative noise in ship; with other cues properly set, the listeners perceived 'gray' when the duration of the fricative noise (in the next syllable) was relatively short, but great when it was relatively long. Thus, the perception of gray could be changed to great by adding fricative noise in the syllable that followed the target word.

Apparently, the listeners in that experiment integrated into a unitary phonetic percept a variety of acoustic cues that stretched over at least two syllables and overlapped completely with cues relevant to other phones. But how does the listener do this? More specifically, how does he know when to stop integrating? Looking at the variety of cases of this type, we conclude that the integration period is marked neither by a temporal criterion (integrate every x msec), nor by an acoustic one (integrate every time a particular kind of sound is heard). Rather, the integration seems to occur over any stretch of the signal that contains the acoustic consequences of just those articulatory maneuvers that are the peripheral reflections of the speaker's intent to produce a particular phonetic segment. We must wonder, then, how the listener delimits the proper span over which to integrate, in what form he holds the pre-integrated cues, and what he does while waiting.

Consider now, though briefly, how analogous this is to what happens in the decoding of syntax. Surely, the meaning of a syntactic structure (e.g., a sentence) cannot be had except as the listener takes account of the words the structure comprises. As in the phonetic case, the size of this structure is not defined by a temporal criterion, nor by an acoustic one. Rather, it appears to be any number of words that are relevant to the

syntactic structure, and that depends, in turn, on the nature of the message the speaker means to convey. Here, too, then, we must wonder how the listener knows when the structure is complete, in which form he holds the words pending completion, and what he does while waiting.

#### References

- Bailey, P.J., Q. Summerfield, and M. Dorman (1977): "On the identification of sine-wave analogues of certain speech sounds", Haskins Laboratories Status Report on Speech Research, SR-51/52, 1-25.
- Bailey, P.J. and Q. Summerfield (1978): "Some observations on the perception of [s] + stop clusters", Haskins Laboratories Status Report on Speech Research, SR-53, Vol. 2, 25-60.
- Cooper, F.S. (1963): "Speech from stored data", 1963 IEEE International Convention Record, Part 7, p. 139.
- Dorman, M., L. Raphael, and A.M. Liberman (1978): "Some experiments on the sound of silence in phonetic perception", (submitted for publication).
- Fant, G. (1973): "Descriptive analysis of the acoustic aspects of speech", Speech Sounds and Features, Ch. 2, 25-6. (Article based on a paper by Fant presented at a Wenner-Gren Foundation Research Symposium held at Burg Wartenstein, Austria, 1960, which appeared originally in Logos, 5, 3-17 (1962).)
- Isenberg, D. and A.M. Liberman (1978): "Speech and nonspeech percepts from the same sound", JASA 64, Suppl. No. 1, J20.
- Lane, H.L. and B.A. Schneider (1963): "Discriminative control of concurrent responses by the intensity, duration and relative onset time of auditory stimuli", unpublished report, Behavior Analysis Laboratory, University of Michigan.
- Liberman, A.M. and M. Studdert-Kennedy (1977): "Phonetic perception", in Handbook of Sensory Physiology, Vol. VIII, "Perception." ed. by R. Held, H. Leibowitz, and H.L. Teuber; Heidelberg: Springer-Verlag, Inc.
- Mattingly, I.G., A.M. Liberman, A.K. Syrdal, and T. Halwes (1971): "Discrimination in speech and nonspeech modes", Cogn. Psych. 2, 131-157.
- Rand, T.C. (1974): "Dichotic release from masking for speech", JASA 55, 678-680.
- Repp, B.H., A.M. Liberman, T. Eccardt, and D. Pesetsky (1978): "Perceptual integration of temporal cues for stop, fricative, and affricate manner". J. Exp. Psych.: Human Perception and Performance (in press).

## ISSUES IN SPEECH PERCEPTION

Dominic W. Massaro, Department of Psychology  
University of Wisconsin, Madison, Wisconsin, 53706, USA

My goal in the present paper is to address what I believe to be some important issues in the study of perception of speech and nonspeech sounds. The issues are discussed in the framework of binary contrasts. The binary framework was deemed appropriate because of both linguistic precedent and limited psychological capacity. Some hierarchical organization of the issues is probably optimal but I have been reluctant to provide one; the reader can sort, add to, delete, and order the issues as she or he chooses.

Templates versus features

Speech sounds may be gestalt units that cannot be further analyzed or reduced in terms of other attributes. If speech consisted of a sequence of indivisible sounds, then speech analysis would be limited to some variation of a template matching scheme. For successful analysis, an additional template would be needed for every unique speech sound. Although this possibility may be linguistically and psychologically correct, it leaves the student very little to do beyond a general recording and tabulation.

Not only does the template matching scheme leave time on the student's hands, it is not very appealing to those of us who wish to impose simplicity and order upon Mother Nature (or Mother Tongue). Luckily, Jakobson and his colleagues of the Prague school successfully argued that phoneme units could in fact be further analyzed in terms of distinctive features that represent similarities and differences with respect to other phonemes. Given this theoretical perspective, it follows naturally that all of the phonemes of a language can be characterized in terms of a set of distinctive features. Feature analysis is appealing because it allows the units to be subjected to a more abstract classification.

Feature analysis is also preferred over template matching in the study of perception. Template matching schemes would not illuminate any perceived similarities or differences among speech sounds. The applicability of feature analysis proves useful in understanding the findings that two sounds are perceived as similar to one another or are in fact confused with one another to the extent they share the same features. Independent evidence for

features comes from well-known neurophysiological findings that individual cells in the cortex respond selectively to a class of stimuli that share a particular property, such as the direction of the frequency change in a sound. Feature analysis is a worthwhile enterprise as long as we sometimes remind ourselves that it must stop somewhere. When a set of descriptors is no longer analyzable we are left with miniature templates.

Binary versus continuous features

Although it is not unreasonable to describe a speech sound in terms of the degree to which a feature is present in the sound, Jakobson made the important assumption that distinctive features<sup>1</sup> were binary in that each feature is either present or absent in all-or-none fashion. Jakobson argued that "the dichotomous scale is superimposed by language upon the sound matter." The idea of binary features is appealing in terms of parsimony but most importantly in terms of ease of classification. The integration of binary information from two or more feature dimensions requires only logical conjunction of pluses and minuses. The elegance of binary classification is probably responsible for what might sometimes be viewed as an excessive observance of the principle.

In terms of speech perception, it seems more reasonable to assume that the listener has information about the degree to which each feature is present in the speech sound. This assumption of continuous rather than all-or-none featural information contrasts with the traditional view of binary features in linguistic theory. More recently, Chomsky and Halle and Ladefoged have allowed a multi-valued representation of featural information at the perceptual level. In our model, each feature is evaluated in terms of a fuzzy predicate that specifies the degree to which it is true that the sound has a particular feature. Given the fuzzy information passed on by feature evaluation, it is apparent that the integration of this information across several features is more complex than in traditional all-or-none classificatory schemes. Much of our work has supported the idea that features are combined in

(1) The reader should be reminded that the issue of binary versus continuous features is independent of other issues such as phonetic versus acoustic features. Accordingly, even though some examples are drawn from linguistic analyses, the use of features is intended to be general and not limited to one level of analysis.

terms of a multiplicative rule. This combinatorial process is extremely simple but has the nice consequence that the less ambiguous features carry more weight.

#### Phonetic versus acoustic features

It is readily transparent that the concept of phonetic features has advanced the study of the linguistic classification of speech sounds. Students of speech perception must further inquire, however, whether speech perception is mediated by phonetic and/or acoustic features. The seminal work at Haskins Laboratories using synthetic speech evolved around the assumption that phonetic features were perceptually real. Many experiments were carried out to determine which acoustic properties of speech sounds were responsible for the perceived presence or absence of phonetic features. Given our analysis in the discussion of templates versus features it follows that the acoustic properties of speech sounds could be evaluated in terms of templates or features. If you agree that feature analysis is more desirable, then the speech perception theorist must be concerned with the analysis of speech sounds in terms of acoustic, not just phonetic, features.

#### Single factor versus multifactor experiments

In most experiments, speech sounds are varied along a single relevant dimension and observers are asked to perceive a given contrast between two sounds. For example, in the study of the acoustic features for a voicing contrast, all acoustic properties relevant to the contrast are made relatively natural except one, such as voice onset time, and this property dimension is varied through a continuum of values. Very few experiments independently vary more than one property within a particular experiment. The few exceptions in the early literature essentially reduced the data analysis to single-property experiments. In our work we utilize factorial designs and functional measurement techniques to study how acoustic features are evaluated and integrated together. With this procedure, two or more acoustic dimensions are independently varied so that all combinations of the values of one property are paired with all combinations of the values of another property. This design allows a direct assessment of how the acoustic features are evaluated and integrated together in speech perception.

#### Independent versus dependent features

This issue centers around whether the value for a given feature is modified by the value of another feature. Some support for featural independence was provided by studies demonstrating that separate sets of acoustic properties were relevant for perception of different contrasts. However, this result does not necessarily rule out the possibility that the perception of one contrast is dependent on the perception of another. Nonindependence has been proposed to account for the observed shifts in a voicing-contrast boundary as a function of a contrast in terms of place of articulation. However, these boundary shifts may occur even if each of the features makes independent contributions to the analyses. The observed interaction may result from the manner in which the independent featural information is integrated together. A quantitative model based on this idea has been successful in providing a quantitative account of boundary shifts and, therefore, the shifts do not imply nonindependence of feature evaluation.

#### Phoneme versus syllable units

Speech sounds of phoneme size have proven to be valuable in linguistic analysis. For the student of speech perception, however, it is important to ask what sound units are perceptually real. Although it is not easy to determine the sound units that are functional in speech perception, the question can be addressed simultaneously with the study of acoustic features in speech perception.

In our model, features are evaluated and matched to those features which define units in long-term memory. A unit is represented in long-term memory by a prototype which consists of a list of acoustic features. We assume that perceptual recognition of speech is mediated by vowel, consonant-vowel, or vowel-consonant syllable units in long-term memory. This assumption contrasts with the more commonly accepted notion of phonetic or phonemic prototypes in which phonetic or phonemic decisions mediate speech perception. Although it is only natural to say that a particular acoustic property cues voicing, the perception of the phonetic feature of voicing does not mediate syllable recognition in our model. Experiments that have evaluated the acoustic properties that are responsible for phonetic contrasts ask listeners to distin-



guish among speech segments of, at least, syllable length. These experiments do not necessarily mean that speech perception of the syllables was mediated by the phonetic contrasts defined by the experimenter.

In addition to the problem of the lack of acoustic invariance for some consonants, phoneme units cannot easily account for the finding that the vowel sometimes provides direct acoustic information about the consonant portion of a syllable. Vowel duration has a large effect on the voicing contrast of a vowel-consonant syllable in word-final position. Experimental and theoretical work in our laboratory supports the idea that acoustic features of the vowel portion and consonant portion are perceived independently, integrated together, and evaluated against syllable units in memory.

#### Stimulus versus process descriptions

Researchers are converging on the belief that there exists a plethora of potential acoustic features in speech perception. In contrast to the relatively small number of linguistic distinctive features, the potential candidates for acoustic features seem endless. Faced with this army of potential features, what might be the most valuable tack to take? Rather than attempting to define and catalog the large family of features, it might be more worthwhile to design prototypical experiments to assess how a small number of acoustic features are evaluated and integrated together in speech perception. The goal would be to develop a testable description of the process of speech perception rather than a complete stimulus description of all acoustic features. Needless to say, good judgment on the part of the speech researchers will allow a gradual accumulation of a stimulus description in their quest for understanding speech perception processes.

#### Acoustic versus contextual determinants

Speech perception research has been characterized by the study of speech perception as a function of acoustic changes in speech sounds. The researchers have not denied that other sources of information may also be exploited in perceiving natural speech. Not long after the investigator begins to understand how acoustic features are evaluated and integrated together in speech perception, it becomes necessary to assess how the processes work when

contextual influences are also available. As an example, feature evaluation and integration could be studied as a function of both acoustic changes in the speech signal and contextual constraints in terms of how likely a given sound may occur in a given context. A quantitative description of analogous experiments in reading supports the idea that contextual constraints simply provide an independent source of information exactly analogous to what would be provided by an additional feature.

#### Speech perception versus speech recognition

Upon reflection, it is apparent that speech recognition does not mirror speech perception. I recognize (and classify) two sounds as the same without necessarily perceiving them as identical. I believe that the idea of perceptual constancies has misled researchers in not only areas of visual perception but also in speech. The receding object is recognized as the same object even though the retinal input undergoes drastic changes. But the perception of the object also changes as is easily demonstrated by a little perceptual scrutiny. Following in the behavioristic tradition, researchers usually ask listeners to identify or classify sounds and take performance as an index of perception. Are we asking observers to make the stimulus error as the early introspectionists would claim or are there experimental tasks and performance measures that provide good indices of speech perception? This issue may help illuminate the general area of categorical perception by asking to what extent categorical perception is not categorical perception but simply categorical recognition.

To more directly tap perception, experimenters might employ continuous rather than discrete response alternatives. A discrete judgment may not be sensitive to the continuous changes in perception produced by continuous changes in an acoustic property of the speech sound. As an example, small increases in voice onset time for a velar stop might be perceived as making the sound more like /ki/. However, if the sound is still perceived as more like /gi/ than /ki/, the listener may always respond with /gi/. If the listener's judgments are consistent, the different sounds would be responded to equivalently even though they are perceived as different. By asking the observer to make a judgment on a continuum between the discrete alternatives, the responses may

more directly mirror perception. We have obtained orderly data from observers marking off a line in order to place the percept somewhere between discrete alternatives.

#### Speech perception versus speech understanding

It is easy to forget that speech perception does not necessarily entail speech understanding and that accurate understanding does not demand accurate speech perception. Consider a lexical decision task in which a listener indicates whether each test is a word or a nonword. The nonwords, such as "prust" and "mantiness", are perceived correctly and could be repeated even though no understanding takes place. I don't think that it would be profitable to argue that nonwords are not perceived. Our last noisy party reminds us that a significant amount of speech understanding can occur without perfectly accurate speech perception. In many highly constrained sentence contexts, the listener understands exactly some of the message before he perceives it. In fact, a few recent studies have provided some support for the idea that understanding can actually modify perception. A more convincing demonstration is how the perceived clarity of the words of a song is enhanced when the listener simultaneously reads them. In any case, it is necessary to distinguish between the case in which the listener resolves a piano sound sufficiently to distinguish it from adjacent sounds on the musical scale and the case in which the sound is also identified as middle C.

In our model, perception and understanding occur at two different stages of information processing. The primary recognition process evaluates and integrates acoustic features and outputs a perceptual experience of a speech sound. The secondary recognition process operates on the perceptual information to impose meaning and, therefore, a relatively abstract encoding. Although these are highly analogous processes, they utilize different categories of information in long-term memory and may be influenced by different properties of higher-order contextual constraints.

#### Speech versus nonspeech

It seems appropriate to close with this issue (or nonissue) since it is the topic of this symposium. Although speech represents language and nonspeech does not, it is important to know to what extent perception of speech is analogous to perception of

nonspeech. Does nonspeech perception derive from an evaluation and integration of acoustic features defined with respect to segments of sound? Remarkable parallels between speech and nonspeech have been reported in recent years. Rather than concluding that serious investigators should return to psychophysical studies of nonspeech in order to understand basic auditory processes, it seems more productive to assume that speech offers so much more for experimental study and that the most direct route to an understanding of auditory perception is to be found in the study of speech perception.<sup>2</sup>

#### References

- Derr, M.A. and D.W. Massaro (1978): "The contribution of vowel duration, F<sub>0</sub> contour, and frication duration as cues to the /juz/-/jus/ distinction." WHIPP Report #8.
- Massaro, D.W. (1978): "Letter information and orthographic context in word perception." Technical Report No. 453.
- Massaro, D.W. (1975): (Ed.) Understanding language: An information processing analysis of speech perception, reading, and psycholinguistics. New York: Academic Press.
- Massaro, D.W. and M.M. Cohen (1976): "The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction", JASA, 60, 704-717.
- Massaro, D.W. and M.M. Cohen (1977): "The contribution of voice-onset time and fundamental frequency as cues to the /zi/-/si/ distinction", Perc. Psych. 22, 373-382.
- Massaro, D.W. and G.C. Oden (1978): "Evaluation and integration of acoustic features in speech perception", WHIPP Report #9.
- Oden, G.C. (1978): "Integration of place and voicing information in the identification of synthetic stop consonants," JPh, in press.
- Oden, G.C. and D.W. Massaro (1978): "Integration of featural information in speech perception", Psych. Rev. 85, 172-191.

(2) The research reported in this paper was carried out with the collaboration of Michael M. Cohen, Marcia A. Derr, and Gregg C. Oden and was supported in part by National Institute of Mental Health Grant MH 19399 and in part by the Wisconsin Alumni Research Foundation.



## WHAT TELLS US THAT SPEECH IS SPEECH?

Quentin Summerfield, MRC Institute of Hearing Research, University Medical School, Nottingham, UK, and Peter J. Bailey, Department of Psychology, University of York, York, UK.

Acoustic analysis and perceptual experimentation have suggested that speech sounds are special and distinct from other sounds. First, no obvious one-to-one isomorphism exists between acoustic and phonetic segments, and the latter have been said to be encoded in the former (e.g., Liberman et al., 1967). Secondly, phonetic perception is apparently not fully rationalised by known psycho-acoustic properties of the auditory system. One illustration of this is provided by the experiments described by Dr Dorman (this symposium; see also Bailey, Summerfield and Dorman, 1977), in which the perception of sinewave analogues of speech is shown to depend upon listeners' interpretation of the signals as speechlike or non-speechlike. There is thus some support for the argument that speech perception entails a special decoding process (Liberman & Studdert-Kennedy, 1977). In what follows we shall explore two related questions: what is the nature of the information which might activate such a process, and to what extent should a specification of this information constrain a formulation of the process?

One simple hypothesis could be that speechlikeness is marked by acoustical attributes which, if detected in an initial stage of auditory analysis, direct the signal to a subsequent stage of special phonetic processing. Such attributes would have to be properties of all utterances, and, of necessity, would have to be un-encoded, unlike the contextually mutable segments whose decoding they would trigger. Possible candidates have been considered to be rapid spectral changes (Haggard, 1971) and the onset of periodic excitation (Allen & Haggard, 1977), but their role as 'trigger features' has not been empirically demonstrable. Furthermore, even if elaborated by a variable criterion for the acceptability of a trigger feature, this hypothesis cannot account for the perceptual duality of sinewave analogues of CV syllables. Here the putative trigger features would have to be intrinsic to the information specifying phonetic identity, and so in failing to meet the criterion of invariance would exceed the categorising capabilities of purely auditory analysis. Paradoxically, to detect such trigger features

successfully, the putatively auditory processor would require the properties of a special phonetic decoder.

These considerations call for a re-appraisal of the model in which signals are routed to one type of processor or another on the basis of prior detection of simple acoustical attributes. They suggest that an alternative solution to the problem of distinguishing speech from non-speech sounds could be that phonetic and generalised auditory processing are accorded in parallel to all acoustic inputs. Phenomenal perception would correspond to whichever process achieved a satisfactory analysis. In proposing such a solution, Liberman, Mattingly and Turvey (1972) suggested that a sound is recognised as speech if a phonetic processor succeeds in extracting phonetic features. Thus the acoustic specification of signals as speechlike is conceived as being isomorphic with the acoustic specification of the cues to phonetic elements, and a characterisation of the former would follow inevitably from a characterisation of the latter. We have already noted that speech is considered an intractable perceptual problem, as a result of the non-invariant relationship between acoustic cues and phonetic elements. This contrasts with the more straightforward relationship existing between phonetic elements and articulatory dynamics, which has led to the suggestion that acoustic cues are interpreted with respect to an internalised knowledge of vocal tract behaviour (Stevens & House, 1972; Liberman et al., 1967). A perceptual model of this kind would seem to involve at least two stages: in the first, a sequence of acoustic elements must be segregated and detected; in the second, these elements must be interpreted, presumably to reconstruct the information encoded in the sequential properties of the signal. Knowledge of vocal tract behaviour may assist the first stage, but it governs the second stage. While we have no doubt that speech perception is inextricably tied to the origin of the signal in a vocal tract, we wonder whether a process of fractionation followed by reintegration would best capture the information endowed in the signal by the continuous articulatory flow of a dynamic vocal tract (see also, Bailey & Summerfield, 1978).

It has been the general conclusion of students of perception that distal events and proximal stimulation relate equivocally, and the traditional response to this problem has been to assert that perception is a constructive process mediated by abstract internal

knowledge (see, for instance, Neisser, 1967). This view of perception is currently coming under increasing scrutiny, urging a re-examination of the peculiarities of phonetic perception. Theoretical appraisal (e.g., Turvey, 1977; Shaw & Bransford, 1977) and empirical analysis (e.g., Lee, 1974; Blumstein & Stevens, 1978) suggest that distal events may have a more veridical, if complex, representation in perceptual data than has generally been supposed. Thus it may be profitable to explore the notion that phonetic percepts are not constructed from discrete acoustic elements by the mediation of articulatory knowledge, but rather that they are specified in acoustic dynamics structured by a speech-specific organisation of the vocal apparatus (e.g., Krmpotic, 1959; Fowler et al., in press). The acoustic signal must remain the focus of our concern, given that an unequivocal reconstruction of articulatory dynamics from the acoustic signal is not possible (e.g., Atal et al., 1978).

Implicit in the foregoing is the assumption that information for speechlikeness can be specified at a single level of analysis, for which the most promising popular candidate has been the level of phonetic processing. This is a necessary view, given that listeners can describe as speechlike even highly schematic analogues of speech sounds, provided they permit a phonetic interpretation. However, the notion that speechlikeness is specified only in the information for phonetic elements is insufficient to account for the certainty and immediacy with which naive listeners can identify utterances in an unfamiliar language as human speech. In recognising as speech snatches of foreign languages heard, for instance, when tuning a radio receiver, we are presumably attending to information of a different kind from that which specifies a sinewave analogue of a CV syllable as speechlike. A particular suggestion by Stevens and House (1972) is that natural speech sounds are characterised by 'certain dynamic or time-varying properties, among which are syllabic intensity fluctuations such as are associated with one of the most fundamental attributes of speech - the vowel-consonant dichotomy' (p. 13). Recent reformulations of the processes underlying speech production (e.g., Fowler et al., in press) provide a means of rationalising the multiplicity of information in a speech signal that specifies it as such. In this view, the speaker progressively organises his articulatory musculature such

that moment-to-moment control need only be exercised over the minimum number of muscle groups during the act of speech production. It is suggested that speech is the concomitant of a set of functionally nested constraints upon the organisation of the vocal apparatus as a whole, so that short-term events like consonantal articulations are nested within longer-term events like the reconfiguration of the vocal tract for successive vowels; these are themselves nested within events of even longer life-spans, such as the speech-specific respiratory synergism (e.g., Lenneberg, 1967). All of these articulatory events are characteristic of speech production, and all endow the speech signal with distinctive dynamic properties to which listeners may be sensitive.

This conceptualisation of speech production, and the type of perceptual attunement it implies, are consistent with a broader view of the development of sensitivity to sound in general. In the natural world, sounds result from the participation of three-dimensional structures in events that occur over time. It is held that the evolution in organisms of sensitivity to vibration in the media that surround them progressed as a developing facility in identifying not just vibration or sound per se, but ecologically relevant events whose concomitants are sounds (see, for instance, Masterson & Diamond, 1973). To a greater or lesser degree, a natural sound is specific to (though not necessarily completely descriptive of) both its particular source, and the particular event in which the source is participating.

Following Turvey and Prindle (1978), therefore, we suggest that the distinction typically made in the laboratory between perception of natural (or even synthetic) speech sounds, and perception of non-natural waveforms like isolated pure or complex tones, should be recast as the distinction between the perception of events and the perception of non-events. In terms of this categorisation, speech perception is a particular instance of event perception, and a general description of the auditory perception of natural events should throw light on the specific problem of perceiving articulatory events. A tentative description could be that the perception of events depends upon the registration of the coherence of information specific to a source and information specific to the transformation wrought upon that source. (See Shaw & Pittenger, 1977.) Thus a preliminary answer to the question of

what is a speech sound could be this: a pattern of sound may be perceived as speech if it cospecifies its source as a human vocal tract participating in a physiologically and phonologically permissible act of articulation. The registration of coherence is analogous to perceiving the solutions to a set of simultaneous equations: the equations provide structure and coherence for the solutions, but no one solution necessarily mediates the attainment of any other. What we understand by coherence may be illustrated further with a visual analogy. When a man runs, he structures light in such a way that both his identity as a man and his act of running are specified optically. When we perceive him running, we detect the coherence of these conjoint specifications; we do not first perceive the actor in order that we may interpret the elements of his act. (For a particularly succinct demonstration of the registration of coherence in the perception of such events, see Johansson, 1974.)

It will be apparent that we lack a formal means of characterising the coherence in speech sounds. Nevertheless, the notion provides us with an appealing informal account of the perceptual strategies adopted by listeners in the experiments on sinewave analogues of speech. When sinewaves were heard as non-speech sounds, we suppose that listeners attended to the elements in the acoustic array but not to their potential organisation. In hearing them as speechlike, on the other hand, they attended both to the acoustic elements and to their organisation, which together specify, albeit in a highly reduced form, a vocal tract undergoing a phonologically permissible act of articulation. Those familiar with R.C. James' photograph, reproduced in Lindsay and Norman (1972, p. 8) will recognise that the foregoing analogously describes the initial perception of the picture as a random array of dark and light areas, and the subsequent perception of a Dalmatian dog walking in dappled sunlight. Both hearing sinewaves as speechlike and seeing the dog are compelling perceptions. It may be that the search for coherence in stimulus information is a general goal of perceptual systems, guided and rewarded by the attainment of clarity (Woodworth, 1947; Gibson, 1969). We note that when listeners began to hear sinewaves as speechlike, their identification functions became more consistent and more categorical.

In summary, we are suggesting that the achievement of speech articulation is to present the information for speech perception unequivocally in the surrounding media. The acoustic signal is clearly the most important vehicle for speech, but we acknowledge also the perceptual importance of the speech-specific optical concomitants of articulatory events (e.g., Miller & Nicely, 1955; see Erber, 1975, for a review). Progress beyond the phenomenological interest of demonstrations such as the perceptual duality of sine-wave analogues of CV syllables requires the development of a vocabulary with which to describe how articulatory events structure sound and light in perceptually accessible ways. The mathematics of this description will be complex. Nevertheless, we are encouraged that optical invariants supporting the visual perception of aspects of one human activity, locomotion, have been formally described (Lee, 1974; Cutting et al., 1978). The rebirth of articulatory synthesis for perceptual experimentation (Mermelstein & Rubin, 1978; cf., Haggard, in press) is one precursor of the attainment of a similar specification of the optical and acoustical invariants supporting the perception of speech articulation: that is, to specify what it is that tells us that speech is speech.

#### References

- Allen, J. and M.P. Haggard (1977): "Perception of voicing and place features in whispered speech: a dichotic choice analysis", Perc. Psych. 21, 315-322.
- Atal, B.S., J.J. Chang, M.V. Mathews, and J.W. Turkey (1978): "Inversion of articulatory to acoustic transformation in the vocal tract by a computer sorting technique", JASA 63, 1535-1555.
- Bailey, P.J. and A.Q. Summerfield (1978): "Some observations on the perception of [s]+stop clusters", Haskins Laboratories Status Report on Speech Research SR53 (2), 25-60.
- Bailey, P.J., A.Q. Summerfield, and M.F. Dorman (1977): "On the identification of sine-wave analogues of certain speech sounds", Haskins Laboratories Status Report on Speech Research SR51-52, 1-25.
- Blumstein, S.E. and K.N. Stevens (1977): "Acoustic invariance for place of articulation in stops and nasals across syllable contexts", JASA 62, S26(A).
- Cutting, J.E., D.R. Proffitt, and L.T. Kozlowski (1978): "A bio-mechanical invariant for gait perception", J. Exp. Psych: HPP 4, 357-372.
- Erber, N.P. (1975): "Audio-visual perception of speech", JSHD 40, 481-492.
- Fowler, C.A., P. Rubin, R.E. Remez, and M.T. Turvey (in press): "Implication for speech production of a general theory of action", in Language production, B. Butterworth (ed.), New York: Academic Press.

- Gibson, E.J. (1969): Principles of perceptual learning and development, New York: Appleton.
- Haggard, M.P. (1971): "Encoding and the REA for speech signals", Quart. J. Exp. Psych. 23, 34-45.
- Haggard, M.P. (in press): "Experience and perspectives in articulatory synthesis", in Frontiers of speech communication research, B. Lindblom and S. Ohman (eds.), London: Academic Press.
- Johansson, G. (1974): "Projective transformations as determining visual space perception", in Essays in honor of J.J. Gibson, R.B. MacLeod and H.L. Pick (eds.), 117-138, Ithaca: Cornell University Press.
- Krmpotic, J. (1959): "Donnés anatomiques et histologiques relatives aux effecteurs laryngo-pharyngo-buccaux", Revue Lar. Otol. Rhinol. 80, 829-848.
- Lee, D.N. (1974): "Visual information during locomotion", in Essays in honor of J.J. Gibson, R.B. MacLeod and H.L. Pick (eds.), 250-267, Ithaca: Cornell University Press.
- Lenneberg, E.H. (1967): Biological foundations of language, New York: Wiley.
- Lieberman, A.M., F.S. Cooper, D.P. Shankweiler, and M. Studdert-Kennedy (1967): "Perception of the speech code", Psych. Rev. 74, 431-461.
- Lieberman, A.M., I.G. Mattingly, and M.T. Turvey (1972): "Language codes and memory codes", in Coding processes in human memory, A.W. Melton and E. Martin (eds.), 307-334, New York: Winston.
- Lieberman, A.M. and M. Studdert-Kennedy (1977): "Phonetic perception", Haskins Laboratories Status Report on Speech Research SR50, 21-60. (To appear in Handbook of Sensory Physiology, Vol. VIII, "Perception", R. Held, H. Leibowitz, and H-L. Teuber (eds.), Heidelberg: Springer Verlag.
- Lindsay, P.H. and D.A. Norman (1972): Human information processing: An introduction to psychology, New York: Academic Press.
- Masterson, B. and I.T. Diamond (1973): "Hearing: central neural mechanisms", in Handbook of perception, Vol. III, Biology of Perceptual Systems, E.C. Carterette and M.P. Friedman (eds.), 408-448, New York: Academic Press.
- Mermelstein, P. and P. Rubin (1978): "Articulatory synthesis - a tool for the perceptual evaluation of articulatory gestures", Haskins Laboratories Status Report on Speech Research SR53 (1), 1-11.
- Miller, G.A. and P.E. Nicely (1955): "An analysis of perceptual confusions among some English consonants", JASA 27, 338-352.
- Neisser, U. (1967): Cognitive psychology, New York: Appleton.
- Shaw, R. and J. Pittenger (1977): "Perceiving the face of change in changing faces: implications for a theory of object perception", in Perceiving, acting and knowing, R. Shaw and J. Bransford (eds.), 103-132, Hillsdale, N.J.: Erlbaum.
- Shaw, R. and J. Bransford (1977): "Psychological approaches to the problem of knowledge", in Perceiving, acting and knowing, R. Shaw and J. Bransford (eds.), 1-39, Hillsdale, N.J.: Erlbaum.

- Stevens, K.N. and A.S. House (1972): "Speech perception", in Foundations of modern auditory theory, Vol. II, J.V. Tobias (ed.), 1-62, New York: Academic Press.
- Turvey, M.T. (1977): "Contrasting orientations to the theory of visual information processing", Psych. Rev. 84, 67-88.
- Turvey, M.T. and S.S. Prindle (1978): "Modes of perceiving: abstracts, comments and notes", in Modes of perceiving and processing information, H.L. Pick and E. Saltzman (eds.), 205-224, Hillsdale, N.J.: Erlbaum.
- Woodworth, R.S. (1947): "Reinforcement of perception", AJPs 60, 119-124.

THE PERCEPTION OF CHINESE SPEECH SOUNDS IN MASKING NOISE  
AND FREQUENCY DISTORTION

Tze-Wei Pao and Yung-Tzue Wei, Acoustical Institute of Nanking  
University

Intelligibility tests of Chinese speech sounds were run under five masking conditions, namely white noise, pink noise, speech noise, meaningful speech interference, and reverberation masking in an auditorium, as well as in a quiet studio. To simulate the actual communication circumstances, the noise was introduced at input and output ends, respectively. The signal to noise ratios were 5, 0, -5, -10 dB with a fixed speech level about 80 dB at 1m from the loudspeaker. In addition, the speech and noise were processed with high pass, low pass, or band pass filtering except in the reverberation condition. A set of simplified but rather sensitive word lists were used, which were based on varying the initial consonants (initial consonants are more sensitive to masking than are final consonants). The effects of masking and frequency distortion on the perception of individual Chinese speech sounds will be presented in this report.

## THE GOAL OF PHONETICS, ITS UNIFICATION AND APPLICATION

Björn Lindblom, Institute of Linguistics, Stockholm University,  
S-106 91 Stockholm, Sweden

Chairpersons: Dennis B. Fry and Gunnar Fant

In trying to propose a formulation of the goals of phonetics I have begun by asking: (i) What are the goals and the methods of any scientific discipline? How does science in general work? secondly, (ii) What is the traditional subject matter of phonetics? and thirdly, (iii) What are some of the potential practical applications of phonetic knowledge?

Theory, explanation and scientific understanding

How do scientists formulate their understanding of the phenomena that they have chosen to investigate? We find generally that in empirical sciences it is in the form of a theory that such understanding is expressed. Consequently much scientific endeavor is directed towards the construction of theories. Accordingly a fundamental goal also of phonetics is theory construction.

Our first diagram (Fig. 1) is an attempt to illustrate in simplified form some of the components likely to be found in all scientific work such as making quantitative observations, deriving numerical predictions from a theory and inventing a theory. Scientists select a certain set of phenomena that they would like to explain. This set is the explananda in the right, empirical part of the diagram. They devise methods of observation whose output is intended to be facts not artefacts.

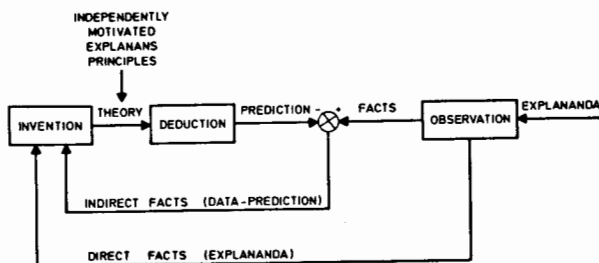


Fig. 1. Some components of scientific investigation.

Moving to the left we find the stage at which facts are compared with predictions or theoretical expectations. This is the point at which the evaluation of a theory begins. Or alternatively, if we have reason to be more confident in our theory than in our methods, it is the point at which we can assess the quality of our measurements. Early in my career as a phonetician I proudly showed Gunnar Fant some spectra that I had produced on the lab spectrograph with what I thought was extreme care so as not to introduce calibration errors etc. Much to my disappointment Gunnar dismissed the data right away and talked about distortion and "spurious formants". Of course he was right. But how could he tell? Later I have realized that the answer is that he looked at the data from the point of his strong theoretical understanding. I find this anecdote instructive since it pinpoints a general problem of research in the several areas of phonetics in which we still lack a powerful theory.

I shall use the term theory to refer to a set of basic laws or principles, on the one hand, and a system of rules on the other. From these basic principles and by means of these rules we deduce mathematically, in a perfectly automatic and formalized way, certain (numerical) consequences representing the predictions of the theory. The job that theories do is to explain. The anatomy of a scientific explanation presents at least the following parts:

1. It presupposes a theory that makes quantitative rather than qualitative statements.
2. It presupposes a theory that is completely formalized and leaves no room for the intelligence and intuition of the person using it.
3. It presupposes a set of explanans principles for which there is ample independent motivation. By independent motivation I mean justification not in terms of the data and the measurements but on external grounds.

In my usage the first two criteria are minimum requirements for an interpretation to qualify a theory. The quality of an explanation appears to be related to two things: the extent to which the theory meets the third condition, that is, has external justification and its scope, i.e., how much data it accomodates.

Summarizing what has been said so far we propose the following tentative definition of scientific understanding: To understand

something scientifically is to be able to recreate one's observations in a quantitative, formalized and explanatory way.

In order to further illustrate these ideas let us move back onto somewhat more familiar ground. Suppose we do an experiment in which listeners are asked to find the best perceptual match between steady-state pairs of synthetic vowels. The reference vowel has four formants. The test vowel has two. The upper formant, the so-called  $F_2'$ , can be varied by the subject. Carlson, Fant and Granström (1970, 1975) did this type of experiment some time ago.

They were able to describe their results in two ways: (i) by means of an empirical formula making  $F_2'$  a function of  $F_2$ ,  $F_3$  and  $F_4$ ; (ii) in terms of an auditory model reflecting the frequency analysis of the auditory periphery.

With respect to numerical accuracy the two descriptions gave almost identical and equally good results. However, when we place these accounts in the context of our previous discussion it becomes clear that only one of them offers an explanation, the one based on the auditory model. Why? Because this description is justified on external grounds. It shows us not only how but also why. It says that the matching behavior of the listeners is simply a consequence of a straightforward cognitive strategy and a phonetic universal: the human auditory system.

The empirical formula explains nothing. It captures certain regularities in the data in a compact and formalized way. It shows how the data came out but provides no clues as to why they came out that way.

Theory and explanation are concepts associated with the ultimate goals of research and it is therefore natural that most of the time we use these terms with restraint. We can name almost any area of phonetics: speech physiology, speech perception, speech development or sound change and we will find that in a certain sense it is true that "we are still at a data gathering stage". Note though that it would be a serious mistake to take this remark to mean that we should abandon all attempts at preliminary theoretical interpretation and model making and concentrate our efforts to the right half of Fig. 1. There are two types of data we need to gather: The facts obtained by direct observation, on the one hand, and the indirect facts represented by the discrepancies between the data and the theoretical model on the other.



Although the predictions may disagree with reality they should nevertheless be regarded as facts, facts about the model. Both the direct and the indirect facts are important sources of information in the creation of models. A good way to learn is to make mistakes in some systematic fashion.

The study of speech sounds: past and present

Phonetics has been traditionally defined as the study of speech sounds. If a deceased colleague of ours active around the turn of the century suddenly rose from the dead and could peep over the shoulders of his modern colleagues he would be unlikely to feel at home in our technologically sophisticated laboratories. However attending conferences and seminars he would no doubt conclude that the major problems to be solved and the questions asked had changed very little. It is instructive to contrast how classical phonetics dealt with the still current fundamental problem of devising a universal phonetic framework for spoken language. This task is essentially two-fold:

First of all, Find a way of describing phonetically an arbitrary utterance of an arbitrary language!

Secondly, Try to represent it in such a way that the description can be reproduced in audible form and with the linguistically relevant features preserved! Here the expression "linguistically relevant features" means the original native accent.

The first problem we can call the analysis or representation problem. The second is that of synthesis.

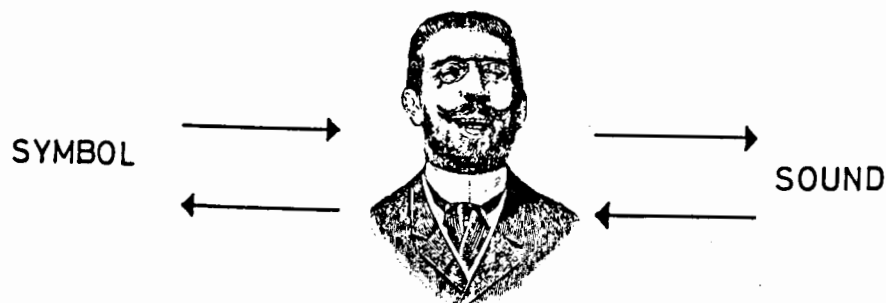


Fig. 2. The solution of classical auditory phonetics to the problem of speech sound specification: the skilled phonetician serving as a human tape-recorder in the "recording" and "playback" of acoustic facts.

The solution of classical phonetics was the concept of the universal phonetic alphabet and the use of highly skilled phoneticians serving as extremely sophisticated tape-recorders in the "recording" and "playback" of acoustic facts. Consider a certain utterance in a given language. Moving to the right in Fig. 2 corresponds to obtaining an answer to the question: What does this utterance, or rather the transcription of it, sound like? Moving to the left: The utterance just spoken by the informant, what is its representation in terms of phonetic symbols?

As we all know this solution of the problem of speech sound specification fails. Its inadequacies cannot be remedied by invoking the important insights contributed later by functional phonemic analysis and distinctive feature frameworks which achieved quantization of the infinite variety of sound and helped define the terms "alphabet" and "universal" more precisely. Nor would it matter if the quest for the ultimate phonetic framework could be brought to a successful close and if suddenly utopian phoneticians emerged capable of using transcription techniques of this type ideally. Why? If science aims at the construction of theories that explain the phenomena under investigation and if contemporary phonetics has the ambition to come of age as a science then it is quite clear why we reject the solution of classical auditory phonetics. This is so because the scientific description of speech sounds must necessarily aim at characterizing explicitly and quantitatively the acoustic events as well as the psychological and physiological processes that speakers and listeners use in generating and interpreting utterances. With the aid of the nimble tongue of the phonetic acrobat classical phonetics succeeds at best in skilfully merely imitating the speech processes of native speakers.

Clearly we must reject the method of impressionistic phonetics because it does not work in practice. Even if it did, it explains nothing: it does not reveal the processes underlying the production and perception of speech sounds. It does not represent a theory in the established sense of this term.

Phoneticians accordingly construe their task of speech sound specification as that of modeling the entire chain of speech behavior in a physiologically, physical and psychologically realistic manner. We thus arrive at the following conclusions: The



traditional subject matter of phonetics is the study of speech sounds; The general goal of scientific disciplines is theory construction and explanation; Consequently the goal of phonetics is to construct a theory of speech sounds; In order to make this theory meet established criteria of explanatory adequacy speech sounds cannot be studied as isolated acoustic events. Speech sounds can only be understood scientifically in terms of the psychological, physiological and physical processes responsible for their generation, on the one hand and with reference to their teleology, that is to their perceptual and communicative purpose on the other. Accordingly the phonetician whose inquiry began at the acoustic level in the domain of speech sounds is today forced to look upstream towards the mind and brain of the speaker and downstream towards the destination of the utterance in the brain and mind of the listener.

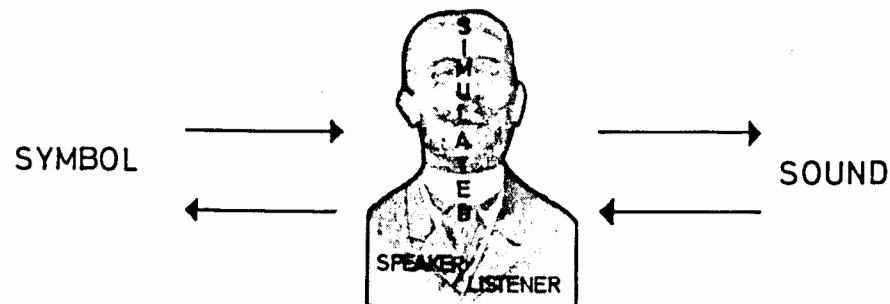


Fig. 3. A goal for modern experimental phonetics: a theory modeling the processes of speaking and listening in an acoustically, physiologically and psychologically realistic manner.

Let us at this point introduce Fig. 3, a slightly modified version of Fig. 2 and recall the phrase we used to summarize our initial discussion of scientific method: To understand something scientifically is to be able to recreate one's observations in a quantitative, formalized and explanatory way.

We can apply this thinking to a larger field of inquiry such as speech production, speech perception or speech development. Or we can apply it to a very restricted set of measurements made in a specific experiment. One very useful measure of our explicit rather than intuitive understanding of the phenomena investigated

is going to be our ability to recreate or simulate them. Needless to say we are in many cases not likely to come close to this goal in the foreseeable future. Nevertheless it provides us with the set of criteria we need to judge the relevance of our short-term efforts.

As we contrast past and present in the historical development of phonetics we see a discipline in the process of transforming from more or less a practical skill or an art into some sort of natural science. This development has yet to be completed but it is undoubtedly an inevitable consequence of: (i) the very nature of the subject matter that we have happened to have chosen; (ii) the natural ambition of any discipline to attain scientific maturity.

We should mention a third factor that has reinforced the present trend namely the prospect of using phonetics for practical purposes. Let me mention a few:

- educational methods and technical aids for the deaf, the hard of hearing, the handicapped and for second-language learners;
  - the diagnosis and treatment of patients with phonetic symptoms including for instance delayed speech development, functional and organic voice disorders, aphasia, hypernasality, dysarthria and stuttering
- as well as
- the automatic analysis and synthesis of speech for various technological purposes.

#### Békésy's mosaic model of scientific progress

In the introductory chapter of his book *Experiments in Hearing*, von Békésy describes his own research in relation to two research strategies: I quote "One, which may be called the theoretical approach, is to formulate the problem in relation to what is already known, to make predictions or extensions on the basis of accepted principles, and then to proceed to test these hypotheses experimentally. Another, which may be called the mosaic approach, takes each problem for itself with little reference to the field in which it lies, and seeks to discover relations and principles that hold within the circumscribed area." Further along in the text: "When in the field of science a great deal of progress has been made and most of the pertinent variables are known, a new problem may most readily be handled by trying to fit it into the

existing framework. When, however, the framework is uncertain and the number of variables is large the mosaic approach is much the easier. Many of the experiments to be described in this book employed the mosaic approach, but when considered in connection with other experiments carried out subsequently by the author and by many other workers in this field they take on a broader meaning and perhaps now may be woven into a more general structure."

Perhaps phonetics is a good example of a field growing like a mosaic. We have profited immensely from technological progress in the form of spectrographs, synthesizers and computers. Clearly such progress has not occurred as a result of premeditated planning on the part of phoneticians but as spin-off effects from adjacent fields with slightly different goals. Recruiting researchers trained in communication engineering, psychology, physiology, mathematics, physics etc. has demonstrably had an extremely vitalizing influence. According to the mosaic model of scientific progress the contents of a field is determined by the questions asked. Eventually a large number of questions will be asked and methods will be developed to answer them. Results will emerge that can be "woven into a more general structure". The lesson taught by the mosaic model thus seems to be: Leave your science alone! Stop worrying about where linguistics and phonetics are going and whether theoretical work is at a standstill or progresses sufficiently fast in response to practical needs etc. I would very much like to accept this advice. But unfortunately the examples that I am going to present to you will lead us in a different direction.

#### Form-based phonetics

When under laboratory conditions Swedish listeners hear the following stimulus:

Tape presentation of left spectrogram of Fig. 4 (next page). Most of them say that they hear the Swedish word hallon beginning with an /h/ and meaning raspberry. What they hear and what you just listened to is in fact the following word simply played backwards<sup>1)</sup>:

Tape presentation of right spectrogram of Fig. 4 (next page). This word means zero. It has the so-called grave accent with an approximately symmetrical rising-falling  $F_0$  contour. The spectrogram to the right thus shows the original recording and to the



Fig. 4. A perceptual paradox: the "nolla-hallon" effect. Left: the Swedish word "nolla" played backwards. Right: the identical word played forwards. Transcriptions indicate perceptual asymmetry.

left we see the backward version. I think you can see that there is a weak expiratory [h]-like noise at the end of [nò:l:a]. Why do our listeners perceive this segment as /h/ when we play the tape backwards but not forwards? One possible interpretation is that this perceptual asymmetry is due to the operation of top-down processes. In other words, you hear in terms of the structure of your native language. Like in many other languages the glottal fricative [h] does not occur in word-final or syllable-final position in Swedish. It does occur in initial position, however. Listeners do not have a sequence \*allon, that is a sequence without the [h] in their lexicon. These facts evidently influence the perception of the acoustic signal in a drastic fashion for the effect is surprisingly strong to native Swedish ears.

The result of this simple tape reversal experiment appears to point to a fundamental principle of linguistic sound analysis: It is language structure and the human ear that determine what is linguistically relevant in the speech wave. The facts of physical phonetics cannot do so no matter how fine-grained we make the analysis. Although initially we rejected the method of classical auditory phonetics we are now paradoxically forced to admit that acoustic-instrumental facts about the behavior we are interested in must be accorded a secondary role in relation to the results of an auditory-functional analysis of sound substance. After all this is very elementary and not very new at all. Think of the notions of segmentation or invariance. Consider for instance the

distinctive feature, the phoneme, the syllable and so forth. All these are linguistic notions in the first place. They have an abstract theoretical status. We bring them with us into our laboratories (and normally we lose them in there before we get out).

Let us consider a statement by Malmberg (1968, 15). In the introduction of A Manual of Phonetics he formulates the role of experimental phonetics in a long-term perspective as follows: "...a combination of a strictly structural approach on the form level with an auditorily based description on the substance level will be the best basis for a scientific analysis of the expression when manifested as sounds. This description has to start by the fundamental analysis, then it must establish in auditory terms the distinctions used for separating phonemic units, and finally, by means of appropriate instruments, find out which acoustic and physiological events correspond to these different units. The interplay between the different sets of phenomena will probably for a long time remain a basic problem in phonetic research." Or take the following statement by Bolinger (1968, 13): "The science of phonetics, whose domain is the sounds of speech, is to linguistics what numismatics is to finance: it makes no difference to a financial transaction what alloys are used in a coin, and it makes no difference to the brain what bits of substance are used as triggers for language."

#### Substance-based phonology

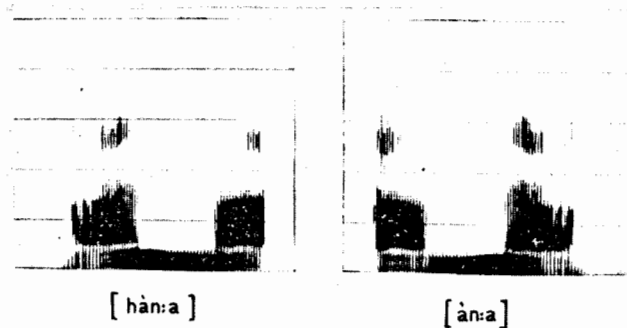


Fig. 5. Left: "Anna" (backwards). Right: the same word (forwards).

Investigating the case of syllable-final [h] further a colleague of mine at Stockholm University Eva Holmberg finds that

this stimulus:

Tape presentation of left spectrogram of Fig. 5 is heard most often as Hanna. What you just heard was the following word played backwards:

Tape presentation of right spectrogram of Fig. 5. Thus subjects clearly hear Hanna rather than Anna<sup>2)</sup> in spite of the fact that both are names and should therefore be in the lexicon of our subjects. Clearly this throws some doubt on our previous interpretation attributing the perceptual asymmetry to language-specific top-down processing. A preliminary look at a large number of languages indicates that the /h/ phoneme tends to be either absent or realized as an [x]-laut or supraglottal fricative in syllable final position. These findings make us favor another hypothesis namely: The parallel between the perceptual asymmetry and the phonological asymmetry is not due to chance. It is due to universal properties of the human speech perception mechanism.<sup>5</sup>

#### The two cultures

The point that I would like to discuss is not whether this hypothesis is correct or not. Rather I have used the case of syllable-final /h/ to demonstrate that this hypothesis cannot be investigated within what Kuhn calls the current "paradigm" of linguistic theory. Given the role that phonetics has played so far in the construction of a theory of language there is no room for a hypothesis of this sort.

What is wrong? Although as linguists we are much concerned with the explanatory adequacy of our descriptions we nevertheless appear to make mistakes of a very elementary nature. In the beginning of our presentation we found that the concept of explanatory theory presupposes that reference is made to principles that are independent of the domain of the observations themselves and that have justification that goes beyond the patterning of the data (cf. vertical arrow at top left of Fig. 1). In common sense terms linguistic behavior presumably arises, both ontogenetically and phylogenetically, as the result of an interplay between

- the functions that language is to subserve;
- biological prerequisites such as brain, nervous system, speech organs, ear, memory mechanisms etc. and
- environmental factors.

Languages thus evolve the way they do because of the body, the mind and the environment. They are the way they are on account of the functions they serve and owing to the properties of both innate and acquired mechanisms of learning, production and perception.

A scientific inquiry conducted along such lines would move our search for basic explanatory principles into the physics and physiology of the brain, nervous system and speech organs, the psychology of the mind and the social dimensions of language use. In other words it would take us right into areas that lie outside linguistics proper and the domains of our primary training and competence.

It might seem as if the strategy that I have been advocating is a reductionist approach to both phonetics and phonology. In other words, adopting this strategy would we then be headed ultimately for "molecular biology" rather than for insights of more primary interest to students of language? My response to this is that there are a host of phenomena for which we do not yet have a very good theoretical understanding. Just to mention a few consider the notions of distinctive feature, segmentation and the syllable and so forth. As long as we cannot treat for instance distinctive features as explananda, as things to be explained, rather than as empirically given primitives - as long as we cannot derive the distinctive features, that is the dimensions of possible phonological contrast, as consequences of constraints on speech communication the reductionist argument has very little force.

The history of phonetics and phonology is the story of two cultures that have always resisted unification. Trubetzkoy (Fischer-Jørgensen 1975, 22) classed phonetics among the natural sciences and assigned phonology to the humanities. The current paradigm of linguistics is aptly termed autonomous linguistics by Derwing (in press). In its context phonetics is a field worth annexing - but for completeness rather than for theoretical relevance.

One cannot help but suspect that autonomous linguistics and the role it assigns to phonetics has developed under the strong influence of educational and administrative constraints and that the program formulated by de Saussure and more recently by Chomsky is a brilliant rationalization of those constraints. If this suspicion is correct - and I truly believe it is - we have reason to examine how we train our linguistics and phonetics students and

how without knowing it we become victims of the irrelevant and conservative influence of how universities are organized in terms of natural sciences, humanities and social sciences and so forth. In that kind of situation leaving one's science alone becomes impossible. However, educational programs can be changed.

#### Summary

We find that the long-term task of phonetics is to contribute towards the construction of a theory of language and language use. This goal is an ambitious undertaking calling for a multiplicity of experimental approaches as well as for theoretical unification.

The question of unification arises in all areas of our field but with particular force as we examine the traditional relationship between phonetics and phonology. We are forced to ask whether phonetics is currently embedded in an intellectual context that is ideally suited for approaching the long-term objectives. Generalizing from the results of a simple but I think instructive perceptual experiment I argue that the answer must be no. The trouble is that the stuff that theories and explanations are made of take us outside the domain of the primary training and competence of phoneticians and linguists. What can be done about this situation? Should we change the goals of phonetics? No, I don't think we can. We are trapped by our choice of subject matter, by scientific method as well as by our obligation to produce knowledge to fields of applied phonetics.

However, phoneticians are not alone in their dissatisfaction with the current paradigm of linguistics. Functionalism has always been alive. We see signs of linguistic research broadening its scope and intensifying research efforts in areas such as sociolinguistics, neurolinguistics, psycholinguistics, language acquisition, sound change, sign language, animal communication and so forth. I think this conference appears to demonstrate a number of such developments which inspire hopes for a new paradigm, a paradigm that views language in a biological perspective and makes it natural and respectable to ask teleological questions - questions that often successfully serve as guidelines for theoretical analysis in other areas of biology (Jacob 1970, Granit 1977) and that in the case of language patterns can be formulated as follows: For what biological and communicative purpose?<sup>4)</sup>

It seems to me that this is the paradigm that phonetic needs.

This is the paradigm in which phonetics will be most effective in contributing towards a better understanding of spoken language. That is a goal worth working for.

#### Conclusion

von Békésy found a close analogy to his research strategies in the field of art. To illustrate the mosaic approach he used a medieval Persian painting with persons and objects represented individually "with little perspective or relation to one another". For the theoretical approach he used a Renaissance woodcut constituting an early attempt to introduce perspective into representation.



Fig. 6. "The Gardener", painting by G. Arcimboldo (1527/30-1593), Skokloster Palace, Sweden.

I was inspired by Békésy to express my final point with the aid of a painting (Fig. 6). I would like to conclude by referring to a portrait by Arcimboldo (1527/30-1593). Let this painting be a symbol of three things: It symbolizes firstly the broad-based, multiple-approach experimental program that we should cultivate, secondly the need for theoretical unification and thirdly the hope that a biological perspective on speech and language will make such unification possible replacing the old paradigms of taxonomy and autonomous anti-functionalism.

#### Acknowledgments

The picture of the Arcimboldo painting was kindly made available to me by Svenska Porträttarkivet, Nationalmuseum, Stockholm.

#### Footnotes

- 1) The "nolla-hallon" effect was discovered about ten years ago by Ulf Ståhlhammar of RIT, Stockholm. I am grateful to him for bringing it to my attention at that time.
- 2) In working on the manuscript of this article I was pleased to hear from G. Heike that the "Anna-Hanna" asymmetry is valid also for German listeners.
- 3) The "nolla-hallon" effect resembles a phenomenon in psychoacoustics known as echo suppression. The sound of a hammer hitting a brick exhibits a certain decay waveform. Comparing backward and forward presentations of this noise one notes a striking asymmetry in that the decay appears much more prominent in the backward playback (Harvard Psychophysics Laboratory: Auditory Demonstration Tapes). There is some recent work on the forward and backward masking of speech-like noise stimuli caused by stationary vowels (Resnick, Weiss and Heinz 1979). This work shows forward masking to be more pronounced than backward masking. It is tempting to assume that the perceptual (and phonotactic?) asymmetry discussed here could be due to asymmetries of temporal masking among other things. However, the literature is somewhat ambiguous as to the direction and magnitude of these masking effects (Holmberg and Gibson 1979).
- 4) Note that I am not advocating some "divine foresight" responsible for order in nature. My model of "purpose" has two components: a "source" generating variation and a "filter" selecting those forms that happen to be compatible with certain "survival" criteria. In language communication the conditions of survival are social and biological in complex interaction.

## 18 PLENARY LECTURE

### References

- Békésy, G. von (1960): "The problems of auditory research", in Experiments in hearing, 3-10, McGraw-Hill.
- Bolinger, D. (1968): Aspects of language, New York: Hartcourt, Brace & World, Inc.
- Carlson, R., B. Granström, and G. Fant (1970): "Some studies concerning perception of isolated vowel", STL-QPSR 2-3, 19-35.
- Carlson, R., G. Fant, and B. Granström (1975): "Two-formant models, pitch and vowel perception", in Auditory analysis and perception of speech, G. Fant and M.A.A. Tatham (eds.), 55-82, London: Academic Press.
- Derwing, B.L. (in press): "Against autonomous linguistics", in Evidence and argumentation in linguistics, T. Perry (ed.), New York: de Gruyter.
- Fischer-Jørgensen, E. (1975): Trends in phonological theory, Copenhagen: Akademisk Forlag.
- Granit, R. (1977): The purposive brain, Cambridge, Mass.: MIT Press.
- Harvard University, Laboratory of Psychophysics (1978): Auditory demonstration tapes.
- Holmberg, E., and A. Gibson (1979): "On the distribution of [h] in the languages of the world", PERILUS I, Dept. of Linguistics, Stockholm University.
- Jacob, F. (1970): La logique du vivant, Paris: Gallimard.
- Malmberg, B. (1968): "Linguistic bases of phonetics", in Manual of phonetics, B. Malmberg (ed.), 1-16, Amsterdam: North Holland.
- Resnick, S.B., M.S. Weiss, and J.M. Heinz (1979): "Masking of filtered noise bursts by synthetic vowels", JASA 66, 674-673.

## DISCUSSION

Victoria Fromkin, Hans Günther Tillmann, and Harry Hollien opened the discussion.

Victoria Fromkin: The question of the boundaries of phonetics and linguistics, or whether such boundaries should be drawn, is an important one. At the Linguistic Society Meeting in Salzburg last week, Charles Fillmore spoke on the question of boundaries, external and internal, in linguistics. The main point was that the goals we have are very often determined by which particular boundaries we set, and where we set them. And what is to one person phonetics may be to someone else garbage. It seems to me that we have to be able and willing to widen our boundaries.

When I first came into the field I was interested in electro-myographic registrations of linguistic units, and there were people who said: "That is not linguistics", and I said: "But linguistics is whatever tells us more about the nature of human language and how language is realized in speech and in perception". - More recently I am interested in the human brain, and I am interested also in mental grammars, and I am even interested in what might go on in one part of the brain as opposed to another, - and people say to me: "That is not linguistics".

I have recently witnessed an experiment with a split-brain patient, whose left hemisphere, and subsequently right hemisphere were anaesthetized. When confronted with pictures of e.g. a mat and a bat, he could not tell them apart, - in fact, he could hardly speak at all, when his left hemisphere was anaesthetized. With the right hemisphere anaesthetized he did very well. Whether one has any quantitative results, whether there is one patient, ten patients, twenty patients, we know that there is something different going on in relation to when a person can tell the difference between mat and bat, and pig and big, and we do not even need to have more than five patients to know that there truly is something qualitatively different in the processing of the linguistic material from the non-linguistic, because when the left side of his brain was anaesthetized, this patient was still able to recognize and sing a song - so there is something special about the linguistic processing going on. Now, of course, we all know that, and what professor Lindblom did reveal is that to understand and to find explanations for this, we must go beyond our perhaps narrow interest and goals, and learn from the physicists, the neurophysiol-

ogists, the neurologists, the psychologists, and gain information wherever we can to try to understand both the nature of language as well as the way we use it in speaking and understanding. It is possible, and I think probable, that there will be certain aspects of human language which we will not find by just these kinds of research. And we will also learn that linguistic systems themselves will give us certain information, in fact raise certain questions as to what some of the rest of us in the laboratory have to seek answers to.

So where I agree with professor Lindblom is that we must go out of our own limited area, seeking help, information, explanations from various disciplines. But at the same time, I think that we should recognize that the autonomous linguists have some very important questions to raise for us to go and do our research on. I think that together we will begin to find out a little bit more about the intricate and complex nature of human language and about those of us who are users of it, the speakers, the hearers, and also the signers and perceivers of sign language, who are deaf.

Hans-Günther Tillmann: Professor Lindblom has drawn our attention to such fundamental and important problems as what it means to say that phoneticians try to develop theories which describe the phonetic facts of speech and language. To further clarify this issue, it could be helpful to turn to two somewhat simpler questions which, on this general metatheoretical level, are somewhat easier to answer: (1) What kinds of facts are given to the phonetician, and (2) what kinds of theories, according to the nature of these facts, can be developed by phoneticians?

(1) It is quite clear that all the facts that phoneticians are concerned with are given by concrete utterances produced by the speakers of a language. It is also quite clear that there are two different types of data to be found in these utterances. In natural circumstances, any such utterance can be perceived by a listener, say a trained phonetician, and hence it can be described symbolically. In this case, the phonetician's data are symbols, and he uses these symbols to refer to certain perceived (or perceivable) facts. Professor Lindblom gave us the two transcriptions [ana] and [hana], and everybody in the audience has learnt under which circumstances each of these transcriptions becomes

true or false - tertium non datur. Quite another type of fact comes into play as soon as we measure co-occurring variations in the physical world. These facts are transphenomenal to ordinary perception, at least in the case of phonetic variations co-occurring with perceivable utterances. If we measure these variations in different areas in and between the brains of the speaker and the listener ('signalphonetisches Band'), we obtain data in the form of time-functions, which in turn can be represented by digital signals. I would like to call special attention to the fact that these two different types of data, i.e. symbols and signals, constitute two different empirical domains for the phonetician - or, as I would like to call it if I could do so in English, two different 'empiries' - which exist separately and logically independently of each other. Perceivable utterances and measurable time-functions co-occur only empirically, yet in an experimentally reproducible (i.e. verifiable) manner.

(2) Given these two different types of data - symbols, representing the category of perceivable events, and signals, representing measurable facts in the physical world - three different types of phonetic theories can be conceived of:

- A phonetic theory can be restricted to symbolic data - we find theories of this kind in phonology - or
- a phonetic theory can be more or less restricted to signals - the causal relations between different time-functions at different points of the physical continuum from the speaker to the listener can be analyzed in order to model the process of transmission of phonetic information from cortex to cortex - or
- a phonetic theory can explicitly try to connect the different facts given by symbols and signals - in this case, the form of a phonetic theory can simply be characterized by saying that the explicanda are primarily given in the first empirical domain of symbols, whereas independent explication can be looked for in the second empirical domain of time-functions.

Phoneticians and linguists are free to formulate and/or invent their explicanda, and they are also free to find theoretical explicata. In this situation, however, I would like to propose that phoneticians (and linguists) should make a virtue of necessity



and let practical applications determine what is to be translated into explicable explicanda. In this case, the solution of practical problems will be the best test to decide whether phoneticians have succeeded in finding a useful explicatum or not.

Harry Hollien: The first question that we have to ask ourselves seriously is: "Are we a discipline? Or are we simply a part of a more important discipline, whether it is linguistics or engineering or speech pathology, or some areas such as these?" - If we do decide that phonetics is a discipline, the second question we have to ask ourselves is: "Can we define it? Can we define its goals, its boundaries, its nature, in such a way that we can articulate this to other disciplines, and is there a cohesion within our field?" And since we represent different nationalities, different philosophies, different backgrounds, different orientations, different fractionalizations, we also have to deal with the third question: "How do we deal with each other, and develop mechanisms, procedures, processes by which to solve fundamentally the disagreements which we have within our field?"

Björn Lindblom: I think professor Hollien is doing it backwards. One begins by raising questions - that is how fields develop, that is how they grow. And: if phonetics is a discipline? I do not think it matters. We are interested in studying speech processes, interested in studying language, and that is where it all begins. And what you are talking about are some administrative, political problems that should be secondary.

I find myself in agreement with professor Fromkin and professor Tillmann. I wish that professor Fromkin would be a little more impatient with the autonomous linguistics paradigm, because it has such a radical influence on what we are doing.

Antti Sovijärvi gave examples of Finnish words which, when played backwards, are perceived by Finnish listeners in accordance with the syllable structure of Finnish.

Gunnar Fant pointed to the fact that there is a physiological explanation for the post-vocalic aspiration in open syllables, i.e. the glottis opens gradually, just like in an h-sound.

Henrik Birnbaum: In Björn Lindblom's initial chart (fig. 1) I was slightly disturbed by the terminology. He used the term 'indirect facts' for data prediction. But I do not think we can talk about fact in any sense here. We can talk, at best, about

hypotheses. I therefore do not think that there is any parallelism between the facts that we are asked to explain and the data predictions that we make, based on partial knowledge. - I think models are supposed to replicate something that we put into the abstract, and I think that what we have as 'indirect facts' in Björn Lindblom's chart, the data predictions, are part of a model, and models are never facts until they are proven beyond doubt correct, - so I would prefer the term hypotheses or partial hypotheses.

If 'autonomous' is understood in a broader way, and not in the narrow sense in which it was used in standard TG grammar, then of course autonomous linguistics, and within that autonomous phonetics, is a discipline. It does not mean that we should cut out all the neighbouring disciplines, however. I also would like to remind you that not only de Saussure and Chomsky would use this term, but Louis Hjelmslev spoke specifically about language as a structure sui generis. Language is a structure sui generis and not a replica of something else. We restructure reality in terms of the system we use.

Fred Peng: I want to ask professor Lindblom if he means that all people, regardless of linguistic or cultural background, hear more or less the same h initially, not heard at the end of the word. - Perceptual asymmetry is not limited to the auditory channel, it is also found in the visual and tactile modes, and I think that the environment, or context, has something to do with what you hear or do not hear, and the brain has sufficient plasticity to enable us to ignore what is not relevant to our background.

Björn Lindblom: We do not deny that listeners of different language background might have different perceptions, depending on their differences in top-down processing, conditioned by their native languages, but we do find parallels in the responses of our Swedish listeners and in the distribution of /h/-phonemes across the languages of the world. And thus we wonder if final /h/'s are not disfavoured because of some kind of auditory asymmetry that we all share. We are not denying that you can make use of this phoneme in final position, but it is disfavoured. It is a near universal absence.

Lise Menn: We need adequate descriptions from autonomous linguistics. It may well be that explanations cannot come from

within linguistics, but descriptions must. Early work in both child language and aphasia is, from a modern perspective, a great mess, - a lot of it, because of a lack of an adequate linguistic theory to relate the data to.

Another point: one level of investigation defines and sharpens the questions asked by another level. When you have gathered data for your theory, then you rephrase your questions - and it is a constant interaction between theory and data that is absolutely necessary. It is very easy to get a plethora of data: the problem is to relate it to theory. What you have is junk unless you know what its linguistic significance is.

Eric Keller pointed out the need for more theoretical papers in phonetics, the lack of which he tied up with the problem of educational background, which needs to be very wide if one is to do adequate work in phonetics. Students should be encouraged to acquire also mathematics, neurophysiology, physiology, psychology.

André Rigault suggested that we stick to de Saussure's distinction between substance and form: phonetics analyzes substance, phonology deals with form. He criticized the use of the term 'experimental phonetics' for something which is, properly speaking, 'instrumental phonetics', because doing an experiment involves having control of the phenomena investigated, to modify them at will. But he also felt that proper experimental phonetics ought to have a prominent place in our work, allowing us to verify theoretical models.

Further, phonetics should benefit from the contributions from psychology, linguistics, engineering, etc., but we should avoid the hyper-rationalization which has taken place in medicine, which produces people with a phenomenal education in mathematics, but no practitioners to cure you of your illnesses.

Suzanne Romaine: I would like to object to the attitude which seems to be implied in professor Lindblom's last remark to the effect that a biological emphasis and perspective is what is needed to unify phonetics and to replace the old paradigms of taxonomy and autonomy, because it reflects a tacit acceptance of a Kuhnian notion of so-called normal science and of science as consisting of a succession of so-called paradigms. I think that unity is the last concept that should be applied to any discipline. We can agree about goals without having to agree on how we are

going to pursue them, and I would like to emphasize my agreement with what Victoria Fromkin said, that there are both quantitative and qualitative aspects to our profession. We do not want to be replacing old paradigms so much as to be increasing competition among paradigms. I think that is the only way for science to grow.

Pierre Divenyi: I would like to expand on the role of biology, from the point of view of perceptual phonetics, and say that maybe we should start learning from what our physiologist colleagues do: at the Cambridge meeting of the Acoustical Society of America in June, physiologists reported on experiments where they have measured the response to speech stimuli of various parts of the auditory system, and I think that now that we know at least how certain levels of this system respond, we should maybe cease considering as a stimulus to the phonetic system the string of phones, for instance, or even the acoustic stimulus itself. Maybe we should consider our proximal stimulus, to demonstrate what is happening at various parts of the system. I would tentatively suggest that the explanation for the 'Anna/Hanna' phenomenon shown to us by professor Lindblom may be deduced from what happens in the auditory nerve.

Fritz Winckel pointed to the parallel between natural sciences, linguistics, and art, all being trial and error processes.

Osamu Fujimura: The point I would like to raise is a general matter of how can we choose the correct criteria for selecting one model among several. And particularly, if there are two models at hand which both of them explain the facts equally well. We should probably be very careful about applying a particular set of criteria, because there are many cases where one experiment or situation does not reveal the entire picture of the subject-matter, and I think that for example in the case of the F2' experiments that professor Lindblom mentioned, isolated utterances, vowels, may not be revealing enough for us to be able to conclude in favour of one model over another.

Jørgen Rischel: It is obvious, to me at least, that we need autonomous linguistic research, at least a research which poses linguistic questions and which does not start out from, say, a biological foundation, and at the same time, of course, we need phonetic research. One of the problems today is that people specializing in different fields do not always grasp the implica-

## 26 PLENARY LECTURE

tions of what people in other fields are doing. For example, it is very important to make clear to what extent a particular distinctive feature framework is motivated linguistically, to what extent it is phonetically motivated by, say, empirical physiological and perceptual research, and so on. There is sometimes a danger of a forth and back reinforcement of one's confidence in model construction: for example some linguistic model may serve as the basis for some phonetic experimentation and confirmation of the possibility of finding a phonetic equivalence, and then this may be used by the linguist as a confirmation of his own research. Therefore, we have to be very careful when we publish our results and make explicit whether we are borrowing assumptions which are not within our own paradigm or research.

## REPORT: SPEECH PRODUCTION

(see vol. I, p. 11-56)

Reporter: Peter F. MacNeilage

Co-reporter: Peter Ladefoged

Co-reporter: Masayuki Sawashima

Chairpersons: Antti Sovijärvi and Hiroya Fujisaki

## REPORTER'S ADDITIONAL REMARKS

P. MacNeilage, in his presentation, commented on the question of the control of speech production and the biological basis of speech.

The first comments dealt with the role of feed-back. P. MacNeilage claimed that if one considers how we produce speech under the various postural circumstances, we are forced to conclude that peripheral somatic feed-back plays a virtually continuous role in the control of speech production. It must be a system that can sense at the periphery what the present posture is and that is required to monitor the attempts of the control system to produce speech in any particular posture. We can't be assumed to be infinitely versatile in terms of preprogramming at all postural circumstances. Furthermore, P. MacNeilage pointed out that the concept of normal speech production is perhaps misleading, since most of our work is done in the laboratory with the subjects looking straight ahead and in a fixed position. This is not the normal posture and very little of our work has dealt with postural variations. P. MacNeilage continued by saying that we know very little about how the feed-back works and that we need more information which may perhaps come from people doing research in dentistry. He warned against conclusions drawn from physiological studies of animal limbs, since the human somatic sensory system differs from the animal system in many significant ways. In addition, P. MacNeilage found that the results of experiments where the posture is artificially manipulated, such as in the bite-block studies and in studies where the jaw movement is impaired, support the argument about the necessity for feed-back.

Then P. MacNeilage raised the question: This feed-back is feed-back to what? Among other things he pointed out that it seems necessary in speech production to recognize a multiplicity of levels of organization, some of which are quite accessible to us

and others which are not. But it is nevertheless crucial for us to understand those higher levels if we want to come up with a plausible theory of speech production. In this connection, P. MacNeilage stated that there has to be a distinction between a context sensitive system at a lower level of organization and some kind of context independent entity or set of entities at a higher level, referring among other things to segmental spoonerism. We produce sequences with spoonerisms fluently which means that subsequent to the permutation, the context sensitive control system makes the appropriate adjustments. He noticed that very often spoonerisms involve single segments, and very few can be unequivocally labeled distinctive feature movement type errors, and relatively few involve whole syllables. This means that at least at one level of organization the segmental unit is an extremely important one for speech production.

Before leaving the topic of control, MacNeilage stated that our rather simple algorithms do not account very well for the dynamic aspects of speech production, referring to differences in stress and speaking rate, and to coarticulation. The same speaker can use different strategies in changing the speaking rate, for instance, which also proves that we are dealing with an extremely versatile control system.

Turning now to the question of the biological basis of speech production, P. MacNeilage emphasized - as he does in his paper - that we have very much neglected the study of prelinguistic vocalization in our studies of speech production. This neglect may be due to R. Jakobson's theory of language acquisition which assigned babbling to "external" phonetics. P. MacNeilage claimed that the phonetic forms of early speech with reference are extremely similar or identical in many cases to the babbling forms that immediately preceded them. This means that the same production system that has been working earlier in the proto-language stage is still an extremely important component in early referential speech. P. MacNeilage claimed that babbling begins at a particular time on a particular day. Finally, he stated that babbling is some kind of innate movement control organization that is "there" in relation to speech.

## DISCUSSION

John Ohala and John Laver opened the discussion.

J. Ohala stated that from his point of view one of the very promising and most essential developments in current work on speech production is the large number of models, including various aspects of the articulatory apparatus, which have been developed in the past decade or so. He believes that the rise of model-making is a development of the computer revolution in the laboratory and that it has come of age where we have become familiar with and have used computers to develop models which in many cases are conceptually simple, but which require computationally rather complex activity. Some objections have occasionally been raised against model-making, usually along the lines of: "Well, you have made the model, you have put the properties into it that it has, why can't you figure out what it is supposed to do in advance, why bother with it? It is simply making explicit what you already know or what you assume to be true." In order to parry off this kind of objection, Ohala referred to the Nobel Prize winner H. Simon, who indicates that it may very well be true that in model-making-like abstract logic and didactic logic and so on - the consequences of a particular set of assumptions must naturally follow in an automatic, perfectly regular way. But when our models and the assumptions in them get sufficiently complex, really only God can figure out what the consequences of these assumptions may be. The rest of us have to work them out painstakingly, teasing them out for understanding, and this is why we make models. Furthermore, Ohala pointed out that our models serve a very interesting heuristic purpose in that they tell us what to look for in the data. This was made evident to him in working with an aerodynamic model revealing that if one is going to have production of a fricative or some kind of fricated segment one should not have nasal leakage, obviously because the air flowing out of the nasal cavity would prevent the build-up of the high pressure drop necessary to produce the turbulence. And Ohala asked whether this has phonological consequences. He pointed out that he had never seen any observation of this in the literature, but when he searched for it he was able to come up with a number of examples from sound change and allophonic variation. For example, English has a palatal fricative

as an allophone of /h/ before the palatal glide /j/ in words like Hugh and human. But that same allophone is no longer a fricative if we embed it in a heavily nasalized environment as in the word inhuman. With this example Ohala illustrates how models can tell us what to look for and in that sense even help us to enhance our naturalistically obtained data base.

Then Ohala addressed one comment to Sawashima concerning the vertical tension of the vocal folds. Sawashima said in his co-report that there is no evidence for the existence of any physiological mechanism whereby vertical compression or tensing of the cords could affect  $F_0$ . However, it is well known that the average  $F_0$  of vowels is positively correlated with the "height" of vowels. But, to date, no one has found any significant difference in the degree of muscle activity of the intrinsic laryngeal muscles during the production of various vowels. On the other hand, van den Berg (1955), Shimizu (1960, 1961), and additional workers cited in Žinkin (1968:353) have found that the laryngeal ventricle is larger, both in width and vertical depth, during the production of high vowels such as [i] and [u] - thus showing greater separation between the ventricular folds and the vocal folds - but smaller during the production of low vowels. Also, Luchsinger and Arnold (1965:223) describe a patient with bilateral paralysis of the cricothyroid muscles but who could nevertheless vary  $F_0$  over a few semitones. X-rays revealed no change in the angle of the cricothyroid visor but the whole larynx was higher in the neck during the production of high  $F_0$ . (More detailed arguments for  $F_0$  variation due to vertical tension have been given in Ohala 1972, 1977, 1978.)

#### References

- Berg, J. van den (1955): "On the role of the laryngeal ventricle in voice production", FoL phon. 7, 57-69.
- Luchsinger, R. and G.E. Arnold (1965): Voice-speech-language; clinical communicology: its physiology and pathology. Belmont: Wadsworth.
- Ohala, J. (1972): "How is pitch lowered?", JASA 52, 124.
- Ohala, J. (1977): "Speculations on pitch regulation", Phonetica 34, 310-312.
- Ohala, J. (1978): "The production of tone", in Tone: a linguistic survey, V. Fromkin (ed.), 5-39. New York: Academic Press.
- Shimizu, K. (1960): "On the motions of the vocal cords in phonation studied by means of the high voltage radiograph movies". [In Japanese; English summary]. Oto-Rhino-Lar. Clinic [Zibi-Inko-Ka Rinsyo] 53, 446-461.

- Shimizu, K. (1961): "Experimental studies on movements of the vocal cords during phonation by high voltage radiograph motion pictures", *Studia Phonologica* 1, 111-116.
- Žinkin, N.I. (1968): Mechanisms of speech, The Hague: Mouton.

J. Laver had four points to make about the issues raised in the three reports.

The first of them dealt with methods for estimating the different muscular forces acting on and in the tongue, as in the work of Fujimura, Kakita, and Perkell. He referred to a finding from speech error work based on an experiment to provoke subjects into making the kind of vowel-blend errors that Rulon Wells claimed almost never happen. The structure of the experiment was to push subjects just beyond the comfortable limit of accurate performance of target vowels. Facing them on a screen were two words - for example PEEP and P ARP - and above the two words were two stimulus lights, and the task was to pronounce each word as accurately as possible immediately the associated light came on. The lights were programmed to come in random sequence, with 200 msec duration, with intervals of 200 msec. In this condition, all subjects made vowel errors, two types of diphthongs and one type of monophthong. When PEEP and P ARP were in competition the two diphthong errors were either PAIP or PIAP. Laver proposed the following hypothesis to explain this result. One might imagine that the commands to the relevant muscle systems had a slight difference in the time course such that if the commands for AR preceded those for EE then one got PAIP and if the commands for EE preceded those for AR one got PIAP. But if the commands to the different muscle systems were issued perfectly simultaneously, then the monophthong [ɜ] as in PURP was the result as the mechanically joint product of the action of simultaneously activated different muscle systems. The relevance of this finding to the problem of estimating relative muscle system forces is, that if we look at the interactions of all pairs of vowels, then the "mechanically joint product" position of the intermediate vowel does not necessarily coincide with the geometric mean position between the two target positions. In the competition between PEP and POOP, for example, the intermediate monophthong was [œ] as in [pœ:p], in other words rather closer to the [ɛ] target than to the [u] target, as the lip position also, one might think, was slightly closer to the [ɛ] target

than to the close rounded [u] target. And this is, as far as the tongue is concerned, presumably because the genioglossus muscle has greater muscular force than the muscle system that raises and backs the tongue. Muscle system interactions of this sort in the balanced protagonist-antagonist situation in ordinary speech may well lie at the basis of the notion of "favoured articulatory zones" in the languages of the world. Laver concluded that we have here a very simple experimental paradigm of competition between two targets programmed in random sequence at high speed which can be applied in many areas of speech production and which can tell us perhaps a number of interesting things about the way speech is represented and controlled neuromuscularly.

Secondly, Laver had a comment about Ladefoged's suggested laryngeal parameters of glottal aperture, glottal tension, and glottal length. He pointed out that one aspect of the usefulness of this approach is that the six main modes of phonation - modal voice (Hollien's term), falsetto, creak, whisper, breathiness, and harshness - all have different specifications on these three parameters. And therefore, an explanatory basis is provided for the mutual compatibility or incompatibility between these six phonatory modes. It means that breathiness and harshness, for instance, are ruled out by that model as mutually incompatible, as they are in real life, because they need very different values on the glottal aperture and the glottal tension parameters.

Laver's third point concerned the habitual mode of phonation adopted by an individual which he found was an excellent example of a muscular setting (Honikman's term). The notion of a setting is extendable beyond the larynx to habitual adjustments of the supralaryngeal tract as well. We are all familiar with people using a particular long-term muscular adjustment of the supralaryngeal tract as part of their habitual voice quality. For example people who raise their larynxes and keep them raised throughout speech, people who have a tendency to maintain the lips protruded, qualities which characterize particular speech communities like velarization that one hears in the speech of Liverpool, and lastly habitual nasalization common among RP-speakers. The nice thing about muscular settings, in the context of MacNeilage's report, is that they furnish an excellent example of the Action Theory concept of co-ordinative structures, tuned to a long-term bias on segmental articulation - just like habitual gait.

The last point dealt with the problem of neuromuscular programming, when it is not just a matter of programming a sequence of segments as such, but rather of programming at least a triple layer of commands. Laver stated that if voice quality has a phonetic component which demands a particular controllable setting of the vocal tract and the larynx, then one has to take care of the neuromuscular programming for that component. Secondly, superimposed on that phonetic component of voice quality there will be the current tone of voice that the person is using, in other words the paralinguistic layer as well. And thirdly, the segmental and other components of the linguistic strand of speech. Laver concluded that neurolinguistic programming in real speech is at least three times more complex than would be needed for any single-layer control of segmental sequence.

M. Sawashima, responding to Ohala's last point, claimed that he did agree that the up and down movement of the larynx is highly correlated to the  $F_0$  change. But Sawashima found it difficult to explain that the up and down movement of the larynx directly can affect the vocal fold tenseness if we consider the mechanical and structural properties of the larynx. Maybe we can explain it by saying that the up and down movement of the larynx indirectly can provide a change in the longitudinal tension of the vocal folds, which was said many years ago by Sonninen and others. Sawashima concluded that what we want to find is a reliable physiological correlate to the change or control of the vocal fold tension, and in that sense we can't say that the change of the vocal fold tension is caused by the up and down movement of the larynx.

S. Smith drew the attention of the audience to some of his works dealing with the functional dichotomy of the vocal folds (membrane-cushion, cover-body) and which were done before the works made by van Berg and Hirano.

P. Ladefoged presented a series of slides showing the laryngeal behaviour for different voice qualities in a Bushman language. In his co-report Ladefoged pointed out that the laryngeal parameters normally used are completely inadequate for a description of the six voice qualities found in this language. A very interesting finding was that the speakers of the language all had



a thickened interarytenoid muscle, which helps them to produce the ventricular phonation. The bulge seen on the interarytenoid muscle is not genetically controlled, because one of Ladefoged's colleagues has developed a thickened interarytenoid muscle, working with the language.

O. Fujimura had two points to make. The first one dealt with spoonerisms as evidence for the phoneme size segment as the functional unit. He pointed out that no unit whether phoneme, distinctive feature, syllable, or word can freely exchange with another unit in any environment. The facts are more complicated, and there are constraints and contextual conditions that have to be considered. Fujimura found that there is a confusion between the elements for exchange and the environment set up for the exchange of the elements, and he proposed to consider not only one unit for everything, the phoneme for instance, but also larger units as well. Typically, the exchange occurs in syllable initial position, and why is it so if the phoneme is really the functional unit for exchange?

The other comment concerned the vertical movement of the larynx, which Fujimura found is a very interesting phenomenological fact in correlation with pitch control. This is quite useful in finding out what the control signals are for "pitch control" in devoiced portions of speech. He referred to Japanese which has vowel devoicing according to certain contextual conditions. Fiberoptic observations have shown some vertical movement, qualitatively, in relation to the lexical accentual patterns and also to the phrase boundary phenomenon. In the case where the second syllable of the phrase is devoiced and should be high in pitch according to the general rules, some native speakers feel that the second syllable in those devoiced cases is low in pitch. And fiberoptic observations seem to support this feeling in terms of the vertical movement of the larynx.

N. Waterson had some comments concerning the question of babbling as preparation for speech. Early babbling or cooing usually begins spontaneously as a type of unstructured vocalization and is generally mainly vocalic in nature with perhaps a few sounds in the velar and uvular regions. This stage seems to be

non language-specific. But Waterson pointed out that the interactions between the baby and his caretakers play an important role in preparing the baby for linguistic communication.

The vocalic type of vocalization is replaced by more complex vocalizations containing various consonantal sounds, and they become structured and repetitive. This suggests that the baby is developing processing skills which enable him to recognize samenesses and differences in vocal stretches - something that is essential for the development of language. When structured babbling begins, mothers tend to imitate those stretches which seem to them to be similar to their own language, so the baby is encouraged to work on the sounds of the language of the environment. The child is thus prepared for the sounds he will use in his first words.

The protolanguage stage, which usually overlaps with babbling, is generally articulatorily much less complex than what has been achieved in babbling but represents the development of the functional use of vocalizations. When vocalization is first used functionally, the production is very simple as if articulatory complex production and functional use cannot be coped with by the child's processing system at the same time at this early stage. When he has learnt how to use simple vocalizations functionally, he is ready for the use of the more complex production of the actual speech, and the first words soon follow.

B. Lindblom pointed out that the interest in the biological basis of speech, brought up by MacNeilage, is an interest in the most general phonological universal of all, namely in the difficult topic of speech sounds being a subclass of all sounds and gestures. In this context Lindblom had a question for Ladefoged, Fujimura, and Perkell, which had to do with our articulatory modelling: "Why leave out the jaw?" Lindblom had earlier argued that with the aid of the notion of neutral tongue shape and the jaw parameter we can perhaps explain the origin of the distinctive feature open and close. Furthermore, Lindblom referred to some jaw data presented in his symposium report showing how consonants resist coarticulation in the environment of maximally open vowels. He found that this illuminates some of the phonetic background on phonotactic syllable structure, on strength hierarchies, and such abstract notions from phonology.

J.S. Perkell mentioned, responding to Lindblom, that the actual contribution of the structure of the jaw - i.e. the lower teeth - to directly determining the area function is minimal, but that the jaw serves more as a framework for carrying other articulators around and thereby has an indirect influence. Perkell pointed out that we can't answer the question concerning the importance of the jaw without including the jaw in our physiological models.

P. MacNeilage, replying to Fujimura, mentioned that what he really wanted to say was to stress the prominence of the segment assuming that the larger the number of areas that involve a unit the more important it is at a particular stage of the modelling process to which one thinks the areas are relevant. He agreed that one has to take into account many units in the modelling process and that contextual influences are extremely important.

Replying to Waterson, MacNeilage pointed out that by babbling he did not mean cooing but just what he liked to call the canonical form, the open-close alternation with time locking. He found that maybe he disagreed with Waterson about the onset of that stage. MacNeilage was of the opinion that it happens rather suddenly. It is an important point that has to be explored in the light of the role of imitation. If the adults imitate the child's forms but the child's initial forms occur suddenly, then imitation may have a rather minor role in the onset of the phenomenon, even if it may be important in its subsequent development.

MacNeilage concluded by saying that he was impressed with the lack of disagreement that there had been about the speech production aspect of the phonetic discipline. He liked to believe that it is a very healthy sign and that the heat of the argument is related to the state of the knowledge in this area.

P. Ladefoged returned to the problem dealing with the jaw. His evidence to say that one should leave out the jaw is that what is controlled is the vocal tract shape, referring to Lindblom and his colleagues, who have shown quite effectively that we can produce very similar shapes with the jaw in different positions. If we look at mathematical techniques for reducing the amount of variance between a group of speakers we come out with factors that

reflect the cavity shapes and do not reflect the jaw positions. This is another evidence that the jaw has no role to play. But Ladefoged pointed out that it is just so for vowels and that he might have to put the jaw back again for consonants, referring to Lindblom's new jaw data for consonants (cf. vol. II, p. 33-40).

H. Fujisaki mentioned that we have to treat the jaw as an independent motor unit when we are dealing with the dynamics of articulation. When Ladefoged speaks about tongue control it is a combination of independent or dependent control of the jaw and the tongue. The fact that one can produce many speech sounds without moving the jaw does not exclude the fact that the jaw plays an important role in articulation.

N. Waterson, responding to MacNeilage, replied that if he by "sudden" meant over two or three weeks then there was probably no disagreement, but if he meant from one day to another then they did disagree. But she pointed out that there is not quite enough data on babbling to be able to make a categorical statement about it.

MacNeilage admitted that he did not have very much data and that much of it was informal, but it was his impression that it happens virtually from one day to the next.

Fujimura advocated the independent function of the jaw. He referred to his tongue model, which actually includes an independent variable corresponding to the jaw angle. Fujimura found that the jaw has important functions particularly with respect to the inflection of stress patterns referring to some of his jaw data, which show that jaw height does not correlate clearly either positively or negatively with tongue height and it is not random either. He concluded that the jaw constitutes a very important articulatory dimension.

J. Ohala made a comment dealing with the interpretation of speech errors. He did agree with Fujimura's call for caution in the interpretation of speech error data for what they may reveal about units of speech production. He did this with an analogy.

#### 40 REPORT: PRODUCTION

Let us imagine the following domestic accident: a cook stores spices in a spice cabinet in alphabetic order, i.e., cumin is after coriander, and tumeric is after thyme, etc. In reaching for the thyme to add to a dish, the cook accidentally grabs tumeric instead, thus making a culinary analogue of a speech error. The analyst trying to interpret this error would look in vain for any chemical or physical similarity between tumeric and thyme. What is the point of this? Simply that for the purpose of retrieval or general "housekeeping" functions of manipulating the stored units of speech, it is possible that the addresses or labels used bear only an arbitrary relationship to the substance of the units themselves. Ohala concluded that until we have some general idea of how speech is "programmed" he did not think that the data from speech errors can unambiguously rule out features, phonemes, or syllables - or something else - as possible units of production.

## REPORT: SPEECH PERCEPTION

(see vol. I, p. 59-99)

Reporter: Michael Studdert-Kennedy

Co-reporter: Hiroya Fujisaki

Co-reporter: Ludmilla Chistovich

Chairpersons: Antony Cohen and Louis C.W. Pols

## REPORTERS' ADDITIONAL REMARKS

Michael Studdert-Kennedy gave a summary of his report. He mentioned that he might have misunderstood the aim of the work of the Leningrad group to some extent. He had thought that they were looking for phonetic segments in the acoustic signal, i.e. for acoustic segments that would be isomorphic with phonetic segments, but it appears from Ludmilla Chistovich's report that they are in fact looking primarily for acoustic segmentation, which will, e.g. be essential for the estimation of durational events.

Discussing the problem of feature detectors he mentioned that animals that have feature detectors and templates (e.g. the bullfrog and birds) have them because they need them, having to get along very soon after birth without parental help, but that is not the case with the human infant, who has a long period of parental care.

Concerning the problem of perception of sounds by means of an integration of a variety of cues, he emphasized that the idea that these cues may be held together by the underlying gesture should not be understood as a claim for a motor theory of perception, which implies that perception requires reference to the production system. The idea is that you perceive the production gesture directly like you perceive the movement of a hand by means of the light reflected from it. If the hand was moved inside a resonating chamber which had a source exciting it, you might hear the gesture instead of seeing it.

Studdert-Kennedy added a section on cerebral specialization not found in the original report. A written version of this addition is given below:

Cerebral specialization

Nonetheless, opposition between the two modes of lexical

access -- holistic, from "auditory contour", analytic, from phonetic segments -- should not be too sharply drawn. The work of Zaidel (1978a,b) with "split-brain" patients has demonstrated that holistic access is certainly possible. The cerebral hemispheres of such patients have been surgically separated by section of the connecting pathways (corpus callosum) for relief of epileptic seizure. The separation permits an investigator to assess the linguistic capacities of each hemisphere independently. Zaidel (1978 a,b) has shown that the isolated right hemisphere of such a patient, though totally mute, can recognize a sizeable auditory lexicon and has a rudimentary syntax sufficient for understanding phrases of up to three or four words in length. However, it is incapable of identifying nonsense syllables or of performing tasks that call for phonetic analysis, such as recognizing rhyme (cf. Levy, 1974). This phonetic deficit evidently precludes short-term verbal store, thus limiting the right hemisphere's capacity for syntactic analysis of lengthy utterances, and forces organization of language around meaning. Whether we assume a similar, subsidiary organization in the left hemisphere or some process of inter-hemispheric collaboration, it is clear that normal language comprehension could, at least in principle, draw on both holistic and analytic mechanisms.

At the same time, Zaidel's work provides striking support for the hypothesis, originally derived from dichotic studies, that the distinctive linguistic capacity of the left hemisphere is for phonological analysis of auditory pattern (Studdert-Kennedy and Shankweiler, 1970). Further support has come from electroencephalography (Wood, 1975) and, quite recently, from studies of the effects of electrical stimulation during craniotomy (Ojemann and Mateer, 1979). The latter work isolated, in four patients, left frontal, temporal and parietal sites, surrounding the final cortical motor pathway for speech, in which stimulation blocked both sequencing of oro-facial movements and phoneme identification.

This fascinating discovery meshes neatly with a growing body of data and theory that has sought, in recent years, to explain the well-known link between lateralizations for hand control and speech. Semmes (1968) offered a first account of the association by arguing, from a lengthy series of gunshot lesions, that the left hemisphere is focally organized for fine motor control, the right hemisphere diffusely organized for broader control. Subsequently,

Kimura and her associates reported that skilled manual movements (Kimura and Archibald, 1974) and non-verbal oral movements (Mateer and Kimura, 1977) tend to be impaired in cases of non-fluent aphasia. These impairments are specifically for the sequencing of fine motor movements and are consistent with other behavioral evidence that motor control of the hands and of the speech apparatus is vested in related neural centers (Kinsbourne and Hicks, 1979). In fact, Kimura (1976) has proposed that "...the left hemisphere is particularly well adapted, not for symbolic function per se, but for the execution of some categories of motor activity which happened to lend themselves readily to communication" (p. 154). Among these categories we must, incidentally, include those that support the complex "phonological" and morphological processes of manual sign languages, now being discovered by the research of Klima, Bellugi and their colleagues (Klima and Bellugi, 1979).

The drift of all this work is toward a view of the left cerebral hemisphere as the locus of interrelated sensorimotor centers, essential to the development of language, whether spoken or signed. To understanding of the speech sensorimotor system perceptual studies of dichotic listening will doubtless contribute. Indeed, important dichotic studies have recently found evidence for the double dissociation of left and right hemisphere, speech and music, in infants as young as two or three months (Entus, 1977; Glanville, Best and Levenson, 1977). However, dichotic work has not fulfilled its early promise, largely because it has proved extraordinarily difficult to partial out the complex of factors, behavioral and neurological, that determine the degree of observed ear advantage (cf. Studdert-Kennedy, 1975). For the future, we may increasingly rely on instrumental techniques for monitoring brain activity, such as the blood-flow studies of Lassen and his colleagues (Lassen, Ingvar and Skinhøj, 1978), induced reversible lesions by focal cooling (Zaidel, 1978b), improved methods of electroencephalographic analysis, auditory evoked potentials (Molfese, Freeman and Palermo, 1975) and, perhaps infrequently, direct brain stimulation.

#### References

- Abramson, A.S. (1977): "Laryngeal timing in consonant distinctions", Phonetica 34, 295-303.
- Campbell, R. and B. Dodd (in press): "Hearing by eye", Quarterly Journal of Experimental Psychology.

- Entus, A.K. (1977): "Hemispheric asymmetry in processing dichotically presented speech and nonspeech stimuli by infants", in S.J. Segalowitz and P.A. Greber (eds.) Language development and neurological theory, 64-73, New York: Academic Press.
- Glanville, B.B., C.T. Best and R. Levenson (1977): "A cardiac measure of asymmetries in infant auditory perception", Developmental Psychology 13, 54-59.
- Kimura, D. (1976): "The neural basis of language qua gesture", in H. Whitaker and H.A. Whitaker (eds.) Studies in Neurolinguistics (vol. 3), New York: Academic Press.
- Kimura, D. and Y. Archibald (1974): "Motor functions of the left hemisphere", Brain 97, 337-350.
- Kinsbourne, M. and R.E. Hicks (1979): "Mapping cerebral functional space: competition and collaboration in human performance", in M. Kinsbourne (ed.) Asymmetrical function of the brain, 267-273, New York: Cambridge University Press.
- Klima, E.S. and U. Bellugi (1979): The Signs of Language, Cambridge, Mass.: Harvard University Press.
- Lassen, N.A., D.H. Ingvar and E. Skinhøj (1978): "Brain function and blood flow", Scientific American 239, 62-71.
- Levy, J. (1974): "Psychobiological implications of bilateral asymmetry", in S.J. Dimond and J.G. Beaumont (eds.) Hemisphere function in the human brain, London: Elek.
- Martin, J.G. (1972): "Rhythmic (hierarchical) versus serial structure in speech and other behavior", Psychological Review 79, 487-509.
- Mateer, C. and D. Kimura (1977): "Impairment of non-verbal oral movements in aphasia", Brain and Language 4, 262-276.
- Molfese, D.L., R.B. Freeman and D.S. Palermo (1975): "The ontogeny of brain lateralization for speech and nonspeech stimuli", Brain and Language 2, 356-368.
- Nakatani, L.H. and K.D. Dukes (1977): "Locus of segmental cues for word juncture", JASA 62, 714-719.
- Ojemann, G. and C. Mateer (1979): "Human language cortex: localization of memory, syntax and sequential motor-phoneme identification systems", Science 205, 1401-1403.
- Semmes, J. (1968): "Hemispheric specialization: A possible clue to mechanism", Neuropsychologia 6, 11-26.
- Stevens, K.N. and S. Blumstein (1978): "Invariant cues to place of articulation", JASA 64, 1358-1368.
- Studdert-Kennedy, M. (1975): "Two questions", Brain and Language 2, 123-130.
- Studdert-Kennedy, M. and D.P. Shankweiler (1970): "Hemispheric specialization for speech perception", JASA 48, 579-594.

Studdert-Kennedy concluded by quoting Ludmilla Chistovich who as a conclusion of her report writes "We (our group) believe that the only way to describe human perception is to describe not the perception itself but the artificial speech understanding system which is most compatible with the experimental data obtained in speech perception research". He found that this was a very good statement of a heuristic programme, but emphasized that what is required is a constant interplay between the psycho-biological facts of the human behaviour and whatever robotic facsimile the engineers have managed to construct.

Hiroya Fujisaki summarized his report, giving a more detailed account of the first section on categorical perception based on slides illustrating his well-known dual coding model of discrimination. The fact that categorical perception appears in an apparent enhancement of discriminability on the phoneme boundary, and not in a suppression of discriminability within the category, was illustrated by reference to experiments with an r-l continuum presented to American and Japanese listeners. Categorization immediately after the auditory mapping and dominance of categorical perception on comparative judgement seems to be characteristic of the speech mode, but is also found in some cases of non-speech stimuli. Due regard should be paid to disturbances by noise (uncertainty) both in the categorical judgement process and in the retrieval process from the short term memory of timbre. The ability of categorical judgement is based partly on basic physical discreteness, partly on language specific criteria acquired through training in a specific language.

As for the perception of speech in context, Fujisaki emphasized that the importance of context can not be evaluated until we have studied the variability of phonemes in isolation.

Ludmilla Chistovich had been prevented from participating in the congress.

#### DISCUSSION

The discussion was opened by Kenneth Stevens, Sieb Nooteboom and Christopher Darwin.

Kenneth N. Stevens confined his remarks principally to the question of invariance versus non-invariance. It is obvious that when one produces phonetic segments in context, the articulators

have to move from one target to the next, and so the signal is clearly context-dependent. But if you examine the sound in the right way and look at the right places in the sound, you will see much less variability and more invariance for a given distinctive feature both in the context of other features in the same segment and in the context of adjacent segments. Stevens showed slides of the acoustic waveforms of the syllables ba, da, ga, pa, ta, ka. The samples were taken at the onset of the consonants and the spectra had been calculated in a specific way with a specific time window. He pointed out that in labials the gross shape of the spectrum was flat or falling and spread out in frequency. For the alveolars the spectrum was also spread out in frequency, but rising, or acute, and in velars it had a prominent peak in the mid frequency range. One may say there is compactness to the spectrum.

It is possible to devise algorithms or templates that will recognize each of these gross spectrum shapes - and the point is that if one looks at the gross spectrum shapes rather than at the details of where individual peaks are in the spectrum, one does see a considerable amount of invariance. Now, this is a physically measured spectrum with a linear time scale and with fixed bandwidths. What one should really do is to look at a spectrum as it is processed by the auditory system with the appropriate bandwidths and time constants of that system. At some level in the auditory representation that spectrum may well be influenced by what immediately precedes the spectrum. There are already neurophysiological data that would indicate that. The spectra would have to be brought more in line with what we know about psychophysics and the electro-physiology of the auditory system. But even at this acoustical level we see a measure of invariance for stop consonants, as far as place of articulation is concerned.

In this connection Stevens added some remarks on categorical perception. As one moves along the continuum from ka to ta, the auditory system does not treat the physical continuum as though you were moving continuously. As long as the sound has some sort of compact spectral peak it would sound pretty much the same, and it is only when this peak disappears that you will get a sudden change over to a different kind of sound. Stevens would argue that at some level of the auditory system there is some kind of unique response to each of the spectrum types characterized by the gross properties mentioned above.

Where should one look in the signal to find this invariance? Ludmilla Chistovich and Stevens agree that the places in the sound where there are rapid changes are the places which seem to contain a lot of information. If one looks at these places, one sees invariance not only for place of articulation but also for other distinctive features. The formant transitions are acoustic material that links these rapidly changing events with the relatively slowly changing events during the vowel. There is a tendency for a given phonetic feature to have invariant properties. Stevens would argue that the infant comes into the world endowed with mechanisms that are sensitive to these properties. It has a mechanism for classifying sounds, in particular features, as being similar. These relatively invariant primary acoustic properties help to define distinctive features and provide the signal with the kind of properties that enable the infant to learn language. The context-dependent effects which can go along with these primary properties can be used when necessary, perhaps in noisy situations or in rapid speech to supplement the primary cues.

Sieb Nooteboom had no disagreement with the description given by the reporters of the state of the art in speech perception research, but some comments with respect to the state of the art itself.

The underlying or most basic common goal of speech perception research is undoubtedly to understand the structures and processes by which a listener can recover from the acoustic signal what a speaker is saying to him. It is only when we have reached a basic understanding of speech perception in this sense that we can apply the insights gained to phonological explanation, improvement of synthesis by rule, etc. The most important of the processes involved may be labeled recognition. But experimental paradigms in our discipline draw heavily on forced-choice identification, discrimination, similarity judgements, and scaling, none of which studies recognition as a process in itself. In a typical recognition task each stimulus is presented once only and is potentially compared by the subjects with, for example, all possible words or morphemes in the language, whereas in identification stimuli are typically presented more than once and the response set is restricted by the task. With a very few notable exceptions (cf. Goldstein 1977, Marslen-Wilson and Welsh 1978, Cole and Jakimik



1978) recognition is not studied at all. In this respect research on reading, where considerable attention is paid to visual word recognition, is ahead of research on speech perception (Bouwhuis 1979).

Too much attention is focussed on phonemes and phonemic features at the expense of more comprehensive structures, words, morphemes, and prosodic structures, and their communicative function. For a listener to understand what a speaker is saying to him, he must generally recognize meaningful units. Words and morphemes are certainly the most important structures in speech perception. Most investigators seem to believe that once we understand how phonemes are extracted from the signal we can easily explain further linguistic processing. This is hardly true. We do not know whether word recognition is mediated by phoneme extraction, or rather, as recently suggested by Dennis Klatt (1979), by spectral templates, and we will never know until we turn to the study of word recognition. And even if word recognition turns out to be mediated by phoneme extraction, that is certainly not all there is to it (cf. the word completion effect in visual word recognition, Reicher 1969, Bouwhuis 1969).

The even more comprehensive suprasegmental or prosodic structures also contribute in several ways to a listener's recovery of what the speaker wants to say to him. It is a good thing that in recent years researchers have been paying more attention to prosodic structures. Attention has mainly centered around the connection between prosody and syntax, but Nooteboom thinks that two other functions are at least as important in daily speech communication. One is that differences in global pitch level, as well as the presence of normal intonational patterning, appear to increase the intelligibility of speech masked by speech (Brokx 1979). The other, and perhaps most important communicative function of prosody is to signal semantic focus (O'Shaughnessy 1978).

We should acknowledge that phonetics, and especially perceptual phonetics, has reached a stage in which it should not be limited to the study of consonants and vowels. Much is to be gained from widening the scope of the mainstream of our discipline.

#### References

- Bouwhuis, D.G. (1976): Visual Recognition of Words. Unpublished Doctor's Thesis, Catholic University of Nijmegen
- Brokx, J.P.L. (1979): Waargenomen Continuïteit in Spraak: het Belang van Toonhoogte. Unpublished Doctor's Thesis, Eindhoven University of Technology

- Cole, R.A. and J. Jakimik (1978): "Understanding speech: how words are heard", in G. Underwood (ed.) Strategies of Information Processing, Academic Press
- Goldstein, L. (1979): "Perceptual salience of stressed syllables", Chapter II of an Unpublished Doctor's Thesis, University of California Los Angeles, Department of Linguistics
- Klatt, D.H. (1979): "Speech perception: a model of acoustic-phonetic analysis and lexical access", Journal of Phonetics 7, 279-312
- Marslen-Wilson, W.D. and A. Welsh (1978): "Processing interactions and lexical access during word recognition in continuous speech", Cognitive Psychology 10, 29-63
- O'Shaughnessy, D. (1976): Modeling Fundamental Frequency, and its Relationship to Syntax, Semantics, and Phonetics, Unpublished Doctor's Thesis, M.I.T., Cambridge, Massachusetts
- Reicher, G.M. (1969): "Perceptual recognition as a function of meaningfulness of stimulus material", Journal of Experimental Psychology 81, 275-280.

Christopher Darwin started by quoting Ludmilla Chistovich (the same passage that is quoted by Michael Studdert-Kennedy at the end of his report). He concentrated his contribution on a discussion of the relation between computer speech recognition work and the human speech perception in the area of auditory feature extraction and phonetic segment identification.

The engineer does not have to make his system in a psychologically plausible fashion to make it work, but there does seem to be general agreement that speech recognition systems should take account of such relatively peripheral auditory phenomena as critical bands, middle-ear transfer function, growth of loudness and non-simultaneous masking although often the application to speech sounds has to be made on trust rather than on adequate psycho-acoustic data.

Chistovich, rightly, identifies as important the problem of how to represent the input parameters to an acoustic phonetic stage. She points out that theories of phonetic perception are going to be heavily influenced by the materials they have to work with. Thus, if speech understanding programmes are to be serious models of human perception we have to find ways of representing the input signal which are more psychologically plausible and more phonetically germane than a series of categorical labels representing the best, fitting one of a small (100-300) number of static spectral templates.



We have rather little idea what the parameters of an auditory representation should be. Probably it should represent all discriminable differences in the speech signal (taking the most liberal view of "discriminable"), rejecting none of the information to which the listener may need to be sensitive (cf. the work on early visual processing by Marr 1976), but on the other hand the representation must be explicit, organised along those dimensions that are most useful to subsequent processing. It is very different to state explicitly that, for example, there is a formant transition passing between two points in the frequency/time space than simply to represent the signal in a "neural spectrographic" form. The former requires extensive additional processing and important choices about what auditory dimensions to represent. These dimensions must allow not only phonetic classification but also the multitude of para- and non-linguistic decisions that we can make on a speech input, together with all those adjustments for speaker and rate of speech which bedevil recognition algorithms.

One property that a psychologically plausible auditory representation must have is to represent amplitude and spectral change explicitly rather than as a sequence of static events. Two experimental reasons can be given why this should be so:

First, the perceived loudness of a sound depends not only on its intensity but on the changes in intensity that precede and follow it. Jesteadt, Green and Wier (1978) have recently documented this effect which they call the Rawdon-Smith illusion after its co-discoverer (Rawdon-Smith and Grindley, 1935); they find that a rapid rise or fall in intensity is perceptually more salient than a slow change, so that subjects will, under suitable conditions match as equally loud two tones of the same duration and frequency that differ by 13 dB in intensity. Perceptually then, steady-states are (at least partly) defined by their edges, not vice-versa.

Second, the apparent perceptual spectrum of a sound is determined by the changes in spectrum that precede (and perhaps follow) it. Haggard and his colleagues (abstracted in Haggard et al., 1977/8) have shown that a flat spectrum can sound like, for example, [i] if it alternates with a sound whose spectrum is the complement of [i] (having zeroes where [i] has poles).

As well as representing change explicitly, the auditory representation must allow auditory properties to be defined relative to a particular sound source. Silence, for example, is not absolute but rather a property of an assumed source. If a continuous formant pattern is perceptually divided into two assumed speakers by rapid alternations in pitch (Nooteboom, Brokx and de Rooij, 1976) then each speaker appears silent while the other is speaking and, with suitable choice of formant patterns, this perceptually induced silence can cue stop consonants (Darwin and Bethell-Fox, 1977).

Finally, Darwin wanted to make it clear that he finds the interaction between psychological theory and computer algorithm extremely stimulating. It is too easy for someone working with synthetic speech as a tool for investigating human perception to equate the auditory or phonetic dimensions used by the brain with the control parameters of his synthesizer. Trying to write an algorithm to detect, say, voice-onset time is a sobering experience for anyone used to generating beautiful synthetic continua. Algorithms applied to large quantities of natural speech are an invaluable complement to the necessarily restricted psychological experiment.

But if such joint perceptual and computer endeavours are to produce a theory of speech perception rather than a pot-pourri of micro-theories, each concerned with particular phonetic distinctions, we need to be more concerned with the general constraints on speech sounds. What is it that lets us hear as an additional extraneous noise the badly synthesized part of an utterance? Or what allows us to hear speech through a masking pattern that, on a spectrogram, deceives the eye (Lieberman and Studdert-Kennedy, 1978)? The answer for some is in "directly perceiving" the articulation, but we are a long way from being able to write an algorithm that can directly perceive.

#### References

- Darwin, C.J. and C.E. Bethell-Fox (1977): "Pitch continuity and speech source attribution", Journal of Experimental Psychology: Human Perception and Performance 3, 665-672
- Jesteadt, W., D.M. Green and C.C. Wier (1978): "The Rawdon-Smith Illusion", Perception and Psychophysics 23, 244-250
- Haggard, M.P., G. Yates, M. Roberts and Q. Summerfield (1977-8): "Onset and offset spectra in the analysis of complex sounds", Annual Report 1-2, M.R.C. Institute of Hearing Research, Nottingham, U.K., 12-13

- Klatt, D.H. (1977): "Review of the ARPA speech understanding project", JASA 62, 1345-1366
- Klatt, D.H. (1979): "Speech perception: a model of acoustic-phonetic analysis and lexical access", J. Phonetics 7
- Liberman, A.M. and M.G. Studdert-Kennedy (1978): "Phonetic perception", in R. Held, H. Leibowitz and H.L. Tenber (eds.) Handbook of Sensory Physiology VIII, "Perception", Heidelberg Also in Haskins SR-50, 1977, 21-60
- Marr, D. (1976): "Early processing of visual information", Phil. Trans. Roy. Soc. B. 275, 483-524
- Nooteboom, S.G., J.P.L. Brokx and J.J. de Rooij (1976): Contributions of prosody to speech perception, IPO Annual Progress Report 11, 34-54
- Rawdon-Smith, A.F. and G.C. Grindley (1935): "An illusion in the perception of loudness", British Journal of Psychology 26, 191-195.

Dennis Fry expressed his admiration for Michael Studdert-Kennedy's report and for the amount of ingenious experimental work covered by the report. He only wanted, as a supplement, to put forward what he considered to be some brute facts about speech perception seen from the point of view of the acquisition of speech. All reporters mentioned this as an important aspect, but only in passing.

The first fact is that the child always proceeds from the referent to the sound distinction, never the other way about. He is paying attention to something in his environment and that gives him the motive to notice a sound distinction. Therefore this use of acoustic factors probably depends very much on an attentional factor, perhaps more than on the capacities for making these distinctions (cf. Carney and his co-authors).

The second fact is that the child evolves his own acoustic cues. It is essential to remember that every individual is free to evolve his own cues. The only constraint is that they must lead him to the right decision, that is to say to be able to recognize the word or whatever it is that has come in.

This means that the child attempts to learn to deal with the phonetic or perceptual space which is engaging his attention, not the whole phonetic perceptual space, and he starts with very simple cues, expanding the system of cues, that is, developing a larger and larger part of the possible phonetic perceptual space as the different references and the distinction between them make it necessary to do so. - And this whole development goes through re-

ception first. You have to be able to receive, to distinguish, before you begin to produce; there is interaction between reception and production.

Dennis Fry thinks that all this is learnt. The fact that in different languages you get very different modes of dealing with the acoustic input is crucial, and the fact that once you have learnt one language you have difficulty in perceiving distinctions not made in your mother tongue, also shows that these things are learnt. Fry is not convinced of the existence of invariants or of any substratum of universal stuff, perhaps with the exception of the ability to distinguish between silence and sound.

As for the interaction between perception and production we do not keep it sufficiently in mind that every human individual being is hearing a completely unique version of his own sounds. Therefore no human being can make a perfect, and not even a very good match between the sounds he is producing and what he hears from somebody else. It is therefore important that the child develops a cue system which enables him to deal with what comes in. When he sends stuff out, he has only to ensure through his feedback that he is implementing the cues which he is using to listen to somebody else. You have only this amount of match. - Therefore Fry rejected the idea of a motor theory, also in the form that listeners should have to infer something about the vocal tract of the other person. This is not necessary if the whole thing is done on the basis of these cues.

Björn Lindblom showed slides of a distance metric box and of a block diagram of auditory analysis inspired by Manfred Schroeder, which, starting from a harmonic spectrum, converting the frequency scale into a Bark scale, and adding an auditory filter and a masking pattern, leads to two quasi-auditory excitation patterns, a quasi masking pattern and a loudness-density pattern. In accordance with Plomp he thinks that the perceived difference between two static stimuli depends on the area between two curves in the auditory excitation pattern. On this basis he and his co-workers try to explain: (1) the F2' data that have come out of the experiments by Carlson, Granström and Fant (in this respect the results are very positive), (2) Flanagan's difference limen data (which Lennart Nord has had some success in explaining), (3) dynamic events, e.g.: Is a vowel formant target identified better in a

dynamic than in a static context? (Karin Holmgren has found that it is not.) This latter result is not totally in agreement with the point that Darwin made, i.e. that the human speech perception mechanism is primarily sensitive to changes, although Lindblom, generally, agrees completely with this point of view.

Lloyd H. Nakatani agreed that phonetic perception is fundamental to speech perception and that, as Studdert-Kennedy said: "Perhaps all these years of studying C-V syllables have not been wasted after all", but now it is important to concentrate more work on prosody and bring more linguistic facts in. In prosody the cues are complex, and there are great idiolectal differences between talkers. We cannot continue generalizing from the Haskins speech synthesizer to the whole population. In some recent papers in JASA a new technique that attempts to cope with more complex perceptual phenomena has been described.

Dennis Klatt emphasized that you should not set up a dichotomy between phonetic segmentation and the possibility of going directly to larger units, like the word. Both phenomena are well motivated. Phonetic segmentation is supported by the fact that the speech production process manipulates units such as segments, and by the fact that one must have a method for understanding new words. But going directly to the word restricts the phonetic strings to look for and helps solving ambiguities. It also helps to interpret durational cues, because, e.g., stress plays a role. One possibly has to build into our model of the perceptual system kinds of constraints that will make for optimal decisions.

Klatt's second point was that there is no logically necessary connection between feature extraction and phonetic labelling. The features may lead directly to words. One should investigate the feature problem by building very simple models of perception, trying if simple psycho-acoustic distance metrics can be used to make predictions of the sort that are made by phonetic data or not. If not, it points to feature detectors. Probably some of the natural quantal categories will come out of very simple assumptions about the peripheral system and the distance metrics.

The context effects mentioned by Darwin will be troublesome for distance metrics, but this does not prevent a solution. The distance metric is going to be a change-over-time kind of metric.

Osamu Fujimura mentioned a recent study at Bell Laboratories by Marian Macchi treating the role of consonantal transition in perceptual identification of vowels which has been published in *Speech Communication Papers* edited by J.J. Wolf and D.H. Klatt 1979. In contrast to what Strange et al. reported (JASA 60, 1976, p. 213-24), Macchi's result demonstrates that vowels in isolation can give rise to a very high accuracy of identification when appropriate care is exercised concerning dialectal problems and the possible difficulty in orthography (Macchi used rhyming tasks instead). It is possible that dialects vary considerably in the phonetic characteristics of gliding, even for so-called monophthongal vowels in English, and these gliding effects are particularly important in the case of isolated vowels as opposed to syllables ending in a consonant, because the VC transition in the latter case reduces or perceptually obscures such gliding effects.

Dominic W. Massaro: It is recommendable to utilize an information processing approach in speech perception, because the goal of this approach is to delineate the stages of processes that occur between the acoustic stimulus and the meaning in the mind of the observer. It has been found that even at an early stage of processing where you are taking raw feature information and integrating it together it is necessary to incorporate what the listener knows in terms of speech he or she has heard before, in terms of constraints in the language, and in terms of possible words or non-words and so on. So even at this early stage we have to develop models that allow the contribution of higher order processes. Rather than opposing bottom-up and top-down processes; what has to be developed are specific formal models that describe the integration of both sorts of information.

As for features Massaro has found that they are not binary. In fact, listeners have knowledge about the degree to which a feature is present in the speech chain.

Pierre L. Divenyi took up the problem of categorical perception as treated by H. Fujisaki. He found that the problem whether perception, and categorical perception in particular, is articulatorily or auditorily bound is an artificial one. In Fujisaki's second stage there may even enter non-speech auditory events. At the higher stage of perception there is no time for a detailed analysis. Categorical perception is a result of applying an

a priori decision process about what to pick from the signal, and this results simply in discrimination peaks and categories.

Steve Marcus argued that intermediate levels between the acoustical signal and the perceived word are only hypothetical constructs. It appears from split-brain studies that in the right hemisphere word recognition is obtained by an acoustic-lexical mapping system. It would be parsimonious to assume that the left hemisphere used the same system, and that the further possibility of the left hemisphere for segmental analysis would be used for special tasks only, such as CV-recognition, rhyme detection and learning of new words. An intermediate stage seems to be necessitated by current work on the combination of acoustic and visual-articulatory cues (lip reading) in speech perception. It would be interesting to examine whether split brain patients can use lip reading.

Secondly, Marcus argued that there is no empirical justification for assuming a phonemic level. It could also be a continuous real time integration, perhaps using some temporal reference points, which may be purely acoustically determined. The fact that initial phoneme detection times are dependent on factors affecting word recognition speaks against the role of phonemes in perception.

Herbert Pilch. Like Sieb Nooteboom H. Pilch regretted that the study of speech perception has been limited to controlled responses to synthetic stimuli. Our goal must be to understand speech perception in routine communication.

Prosodics signal neither syntax nor sentence meanings, but discourse structuring in the rhetorical sense. Monotonous reading fails to achieve communication, whereas intact prosodic performance can outweigh severe aphasic disturbances in phonemes and syntax.

Routine perception works on the basis not of specific linguistic elements (such as phonemes, syllables, words, sentences) but of total messages. Minimal distinctions may be hard to grasp.

The listener may, however, shift the focus of his perception from the total message to any particular element, i.e. perceive the speech signals as (a) a message, as (b) a linguistic structure, or as (c) noise. In case (a) he may miss the message, in case (b) the structure (cf. H. Pilch: Auditory Phonetics, Word (in print)).

James Pickett: Taking up Studdert-Kennedy's hypothesis that we perceive the speech movements directly, Pickett proposed that we should attempt to set up features of movement (What is moving? where is it going? how is it moving? how is it related to preceding and following movements?) and see where it leads.

Adrian Fourcin: Referring to Dennis Fry's contribution Fourcin confirmed that children do indeed go from the recognition of very simple physical features to levels which are more recondite and varied in the spectral form of the signal. So with the voiced-voiceless opposition you go initially, in the earliest years, from three to five, from a skill of discrimination based on whether voicing is there or not, to a skill based on the onset of the first formant as flat or rising.

Children who are totally deaf can learn to produce clear stress contrasts by means of a visual display of auditorily relevant information. Moreover, by using an auditory pattern approach and giving them an electrical stimulation of the cochlea you can teach totally deaf children to make discriminations based on their pattern knowledge and give them a categorical ability to discriminate which is not at all based on any motor references.

But in order to communicate at a fast rate you have to use a sort of parallel processing technique which is necessarily dependent on your knowledge of coarticulatory constraints.

T.M. Nearey reported that Assmann (cf. vol. I, p. 221) obtained the same results as Marian Macchi (see Fujimura's contribution to the discussion), i.e. a much higher recognition of isolated vowels than should be predicted according to Strange et al., when factors of dialect, orthography, etc. were controlled.

Hiroya Fujisaki emphasized that the role of prosody may be quite language specific. Further, he showed a number of slides illustrating his acoustical and perceptual investigation of Japanese accent.

Michael Studdert-Kennedy concentrated his final remarks on four points:

1. The problem of recognizing dynamic vowels against isolated ones is very complicated. O. Fujimura has showed that centers of vowels extracted from running speech are not readily identified and do need the surrounding formant transitions. Percent correct identifications is probably not the most sensitive measure for that question.

2. Studdert-Kennedy had not attempted to argue that we have no acoustic property detectors. Presumably there is some system within the brain that is able to pick up acoustic properties, but the question is whether there is any grounds for supposing that those property detectors are opponent detecting systems, and whether there is any ground for supposing that they have been adapted for linguistic purposes. In this regard he would rather go with Kenneth Stevens and suppose that language is simply exploiting properties of the auditory system rather than the other way around.

3. In answer to Steve Marcus: To what extent you use auditory contours in listening is an open question. But Studdert-Kennedy would give most of Marcus's data an exactly opposite interpretation. For instance, the fact that phoneme recognition comes after word recognition has nothing to do with perceptual processes, it is a question of experimental tasks and of bringing things into consciousness.

4. Studdert-Kennedy found the data on child language acquisition very important, for instance the work by Boysson-Bardies and by Lise Menn. Another field of research which is highly relevant for the problem of speech perception is that of sign language. Many of the processes of acquisition resemble quite closely the processes of acquisition of spoken language which suggests that what we are dealing with is a very general system that is highly flexible and adaptable to a variety of different circumstances.

REPORT: PHONOLOGY

(see vol. I, p. 103-152)

Reporter: Hans Basbøll

Co-reporter: Stephen Anderson

Co-reporter: Joan Bybee (Hooper)

Chairpersons: William Haas and Kenneth L. Pike

REPORTERS' ADDITIONAL REMARKS

Hans Basbøll: On an abstract level of discussion, it is very hard to disagree with Anderson's claim that one should avoid a priori statements about psychological reality and other linguistic issues, as well as "the arbitrary imposition of restrictive principles which rule out otherwise well-motivated descriptions" (p. 142).<sup>1</sup> I also fully agree with the claim that formal questions just like other scientific questions should be taken seriously.

I am in agreement with the claim that the very fact that part of the traditional field of study cannot be dealt with adequately within a certain framework is not a decisive argument against the use of that framework in other parts of the field. Thus I would suggest that the SPE approach towards markedness, which is considered quite unsatisfactory by both of my fellow reporters, can in principle be used in a rather specific subpart of that subfield of the study of sound structure which it was devised to deal with: namely, to account formally for implicational universals à la Roman Jakobson between sound types. What is outside the scope of the SPE approach towards markedness and similar approaches are other aspects of natural systems and natural segments (like prohibited segments and contrasts, or internal economy) as well as explanation, in any interesting sense, of the relation between phonology and phonetic substance. Such an explanation remains an important task of our discipline, of course.

While I also partly agree that certain efforts of Natural Generative Phonology might be termed reductionist, namely the axiomatization of strong constraints on the form of grammars, I would, on the other hand, suggest that a considerable part of the

-----

1) Pages refer to volume I.

efforts of SPE-phonologies is reductionist in the sense that large amounts of evidence, and thus potential counter evidence, is not taken systematically into consideration. The data considered as evidence is too often limited to a static set of occurring forms as against all the facts which the languages present (cf. p. 142), including those that may be revealed in psycholinguistic experiments, and in studies of language acquisition, language loss, and so on.

What is really at issue are two related fundamental problems: first, the question of predictability and second, the relation between model and reality - in particular: What is the model a model of? and how can it be tested?

I would like to emphasize that in my report I have not stated nor implied nor suggested that the goal of phonology is complete predictability (compare also Labov's variable rules which are probabilistic rather than deterministic). I have said, however, and that evidently is not very new, - that a scientific description should be prognostic in the sense that "it should make predictions (which in principle could be refuted) about something outside the material on the basis of which it was constructed in the first place" (p. 117). That phonology could or should in principle be deterministic is a claim which would hardly be defended by anyone to-day, with the possible exception of a few radical behaviorists. I also think that most linguists would accept the hermeneutic goal of "ex post facto understanding" (p. 140), at least faute de mieux. I certainly also agree that the identification of mutually inconsistent principles may advance our knowledge (for instance the "internal" vs. "external" economy of sound systems according to Martinet), but in such cases our efforts should be directed towards finding constraints on the principles in question to diminish (or better, remove) the field of conflict between them. That a phonological description or theory should be prognostic, on the other hand, is a necessary condition for its being even partly tested for falsifiability, that is for one type of decision on how it relates to "reality".

What the model or theory is a model or theory of is, of course, a vexed question which is closely related to the issue of the reality of phonological descriptions in general, either psychological or sociological. I shall not go into that matter here,

but only briefly remark first that the frequently used phrase 'linguistically significant generalization' may have very different meanings according to the type of reality - if any - ascribed to phonological or other linguistic descriptions; and second, the question of psychological reality is not of the yes-no-type, but there would be a whole scale of possible relations between some internal grammar and an observationally successful model of it (as far as its output is concerned), stretching from a "black box" to a point-to-point-correspondence.

The relation between model and "reality" is of a dialectic nature: The model specifies a number of theoretical constructs, like "natural class" in the "model-internal" sense, defined as a certain set of co-occurring distinctive features, to take just one example. At the same time, real languages present natural classes of segments in the "model-external" sense, that is sets of segments that function as a class in real processes in languages, be it acquisitional, synchronic, diachronic, or whatever. The testing and modification of this part of the model is then a series (generally an infinite one) of steps whereby the sets of segments specified by the "model-internal" and "model-external" natural classes should be brought to coincide, while still respecting all other conditions on the theoretical constructs, such as other types of criteria for the establishment of distinctive features. The model specifies which types of data we should look for, and also which aspects of the data should be considered pertinent and which aspects irrelevant; it must then be independently decided whether the data is in conflict with the model or not.

Now, the point is that this partial testing procedure presupposes that the parts of the model not under consideration for the given purpose must be treated as given for that purpose (as I have said in my report): you cannot test everything at the same time. This is all right if the scientific paradigm within which you work is accepted as basically correct in its main lines, and that is exactly where a clear and fatal division of attitude towards the state of the art occurs, in particular whether the "conceptual richness" of SPE in Anderson's words (p. 136) corresponds to anything outside the model itself. Some people, like my fellow reporter Stephen Anderson, think that SPE represents



"monumental results" (p. 138) and that it is methodologically sound whereas others, including myself, consider SPE - despite its monumental efforts and certain merits - as misguided in quite fundamental respects. I should like to stress once more that both of these two attitudes towards a research paradigm may per se be scientific.

Stephen Anderson: I will focus my attention on the apparent conflict between rationalist and empiricist approaches to sound structure, this being a distinction that I think is at least operationally similar to that raised by Basbøll as the distinction between formal and substance based approaches. This distinction can usefully be approached in terms of the following question: After we have taken into account all those aspects of speech that are associated with more general problems, and which can be approached from outside the domain of language per se, how much is left? Substance based views have typically pursued the possibility that virtually all aspects of language are accessible from one or another more general point of view, and that they can be treated as special cases of the functioning of the articulatory apparatus, of generalized perceptual strategies, of general limitations on memory and processing, and the like. As a result, these researchers have put a great deal of faith and emphasis on the possibility of experimental verification of the details of linguistic structure, for example on the devising of psychological tests to determine on the basis of constructed tasks whether particular proposed phonological rules are psychologically real or not. The substance based linguist takes the absence of such external evidence as establishing a case ex silentio against the proposed analysis as a correct account of language.

The formal approach, on the other hand, has been motivated by the feeling that there are distinct aspects of language which are proper to itself, not studyable necessarily as special cases of other systems. Hence, for the formalists, the absence of direct external accounts for some area of language is not very surprising, or a cause for alarm. This is because this line of reasoning allows specifically for the possibility that among the interacting domains that contribute to the facts of speech, we may find a language faculty which is not indeed reducible to features of other kinds. If so, there is no reason, in principle,

to expect that such a language faculty, if it exists, ought to be directly accessible to inspection in other terms, through constructed psychological experiments of a given kind, for example. The validation of claims of this sort then, would rest not on the establishment of direct links between them and external observables but rather on the inferences that can be drawn from the success, or lack of it, which they achieve in facilitating and revealing regular connections among phenomena, and in uncovering orderliness and coherence within the complexities of languages.

It is important to see that the primary issue between these two views, that of the existence of a specifically linguistic aspect of cognitive structure, not accessible in other terms, could probably never be settled conclusively. One might, of course, establish that a given aspect of linguistic structure is a special case within some more general demand. However, if we construe the proposal that there are aspects of language which are systematically not studyable in such terms, we construe that proposal as an empirical proposition about the nature of language. It is hard to see such a position as other than completely mystical in the extreme. This is, however, not really a matter of empirical fact, but rather a matter of choice of research strategies. Whether or not one ought to limit the terms of linguistic description to elements that can be given an external foundation. As a matter of choosing between research programmes, it seems to me that the claim that all aspects of linguistic structure ought to have some more general basis and ought to be accessible from some other realm, is at least equally mystical, at least in the absence of any such account from any area of linguistic phenomena. The best way to motivate the decision on this issue is to attempt to establish not the correctness but the plausibility of one or the other position. One does this, of course, by demonstrating the ability of this position to provide satisfying and detailed accounts of regularities among the facts of natural languages.

To my mind, the formalist, or as I would prefer to say, the rationalist approach has much the better track record in this regard, though I am sure there are many who will disagree with that. Nonetheless, I hope to have suggested that the choice is by no means an obvious one and in particular, that the formalist pro-



gramme is in no way vitiated, as is sometimes suggested, by its indirect relation to surface facts; that is indeed its essence and its greatest interest.

Joan Bybee Hooper: In the transformational generative tradition a working hypothesis seems to be that if X and Y show some characteristics in common, then they must have the same underlying form, so this produces an emphasis on similarities among elements and has led to a dismissal, occasionally, of surface differences. The results are hypotheses that are untestable because it is always possible to invoke what Botha calls blocking devices, caveats that put hypotheses beyond the surface phonetic facts. This position is exemplified by SPE. The contrary position, which is the one that I accept, requires that linguistic hypotheses be testable (either by comparing them with the surface forms of language or by some kind of experimentation). This is not an a priori constraint on a theory of phonology, it is a different way of approaching facts. Nor is it an attempt to do phonology without an appeal to any abstract entities, because, in fact, all phonology is abstract.

Basbøll expresses the opinion that there is not a big division among these two approaches to phonology. He says in his written report that they share common bases of argumentation and understand each other reasonably well. It seems to me that this is not always the case. There is not a single set of shared assumptions and, in fact, some misunderstanding does ensue. In his paper, Stephen Anderson presents an example from Javanese, intended to falsify the claim that morpholexical rules should apply prior to purely phonological rules. But all we can conclude from the data is that the morphological rule must apply to basic adjectives with round vowels in final position. Only if we assume that lexical representations cannot contain any information that is the output of productive rules does it follow that the morphological rule must apply after the phonological rule. If we do not make such an assumption, the example shows that lexical representations, i.e. the phonological representations relevant for word formation, contain predictable phonetic detail, or to put it another way: the lexical representation has been restructured to contain the output of productive phonetically conditioned processes. The example shows an important difference between the

two approaches: in generative phonology it is assumed that underlying representations are negatively defined by the rules, but I believe that underlying forms and rules can and should be determined independently of one another by examining various types of linguistic evidence and independent or non-structural evidence.

In a paper by Donegan and Stampe in the volume edited by Dinnsen from the Bloomington phonology conference, they characterize a theory of natural phonology by saying: "This is a natural theory in the sense established by Plato in the *Cratylus*, in that it presents language as a natural reflection of the needs, capacities and world of its users, rather than a merely conventional institution. It is a natural theory also in the sense that it is intended to explain its subject matter, to show that it follows naturally from the nature of things. It is not a conventional theory in the sense of the positivist scientific philosophy which has dominated modern linguistics in that it is not intended to describe its subject matter exhaustively and exclusively, i.e. to generate the set of phonologically possible languages." This characterization has two parts: The first one deals with the difference between whether the explanation for linguistic structure will come from general properties of human users of language, or whether it is contained in something that is specifically linguistic and not accessible to verification (although it is not clear to me how this specifically and uniquely linguistic thing is immune to experimental investigation). Secondly, they say that the goal of a natural theory is not to produce exhaustive descriptions of its subject matter. It seems to me that trying to meet the goals of observational and descriptive adequacy has often forced us into making unwarranted theoretical decisions which we may at the time characterize as arbitrary, but in fact then we accept them and never go back to reexamine them; however, such assumptions should be reexamined in view of empirical evidence. Notation is the tool of a theorist and should not be mistaken for the theory itself.

## DISCUSSION

Charles-James N. Bailey, Edmund Gussmann, and Henning Andersen opened the discussion.

Charles-James N. Bailey: Basbøll stresses the role of prediction and explanation. But he does not observe that development is what explains states and their structures; states cannot predict anything but what is in their own scope, and they can explain very little. For minilectal linguists - those who posit idiolects as the object of linguistic investigation and accordingly limit their models to static models - logic suggests that they should give up the goal of exact prediction.

Stephen Anderson's position is quite consistent with his synchronic orientation. He claims that markedness is getting vaguer; but developmental linguistics has been able to define naturalness and markedness quite exactly. Two kinds of dynamic data are relevant for defining the natural and for analysis and description: dynamic changes and comparative patterns (pattern is created by the dynamic principle). With the anticomparative models of minilectal linguistics - phonemes, idiolects, dialects, etc. - the theoretically interesting aspects of linguistics are virtually ruled out, for they demand comparative analysis: naturalness, child language, historical and dialectological linguistics, etc., which are all excluded on principle according to the definitions of phonemes, idiolects, etc. To study development with static tools would be worse than trying to drive a nail with a screwdriver. Since patterns of development are gradient, non-gradient tools are likewise fairly useless. One cannot even describe the morphology of German nasal-stem masculine nouns adequately, for example, with non-gradient models.

Aside from gradience, larger conceptual differences separate the underlying segments of three theories: (1) The classical (taxonomic) phoneme was neither internal-reconstructive nor comparative. (2) The generative phoneme is internal-reconstructive but not comparative. (3) The phoneme is both internal-reconstructive and comparative, or polylectal. Only the latter is valid for development (comparative tasks, including child language acquisition), for theory, and for pedagogy. Development has two sides. One is the inner-linguistic side, where explanations

in phonetology (dynamic phonology) must be sought in phonetics and ultimately in anatomy and bioneurology. The other side is the social side: a development must not only come into existence among children, but must also be adopted by others if it is to survive. Developments due to social or extralinguistic causes may be natural-like, or they may be, and often are, unnatural as in the borrowing of older or of foreign forms, hypercorrect rule-inhibitions, etc. This side of language is only semi-theoretical since many of the relevant conditions are hardly predictable, though creolistics is getting better at predicting changes under different social conditions and with different types of linguistic mixtures. Since Stephen Anderson seems to have a rather negative view toward extralinguistic explanations as well as doubts about some of the explanatory achievements of phonetics, he seems to be skating awfully close to advocating an YROEHT instead of a THEORY; An YROEHT predicteth not; - neither can it explain.

Since it is clear that some linguistic developments are natural and that some are not, and since all languages are mixed and have both of these elements, the immediate goal of linguistics ought to focus on understanding only natural developments and leave the rest for the future.

The abstractness controversy is merely an off-shoot of the really fundamental issue, namely, what are the facts to be analyzed? Our differing views on what is really real affect our views on what data are really relevant to linguistics. If I say that languages have both natural and non-natural phenomena, and you disagree, how could we ever agree on what data are to be admitted or excluded from linguistic analysis?

Even in connection with derivative matters there are several issues of phonetological analysis which are more fundamental than abstractness: There are reasons for believing that instructions from the central nervous system to the articulators are bundled differently in syllable-timed languages and in stress-timed ones, viz. in syllable-sized units and in measures, respectively. One of the deepest issues today is to specify the differences between phonomorphological and morphophonetic (phonetological) rules. Another matter of interest is the fact that the segmental and suprasegmental uses of prosodic features are different: several

rules of English are respectively forwarded and hindered by these different functions of length.

Stephen Anderson takes the wrong view towards different historical developments and their use in the erection of a predictive theory. The difficulties exist only if one excludes the appropriate answer and mechanism: creolizing substrates and superstrates.

If you deal with idiolects, you can always say: "that is your idiolect, not mine", which effectively excludes both proof and replication - and theory. The best way to do linguistics is the way children and adults "do languages", viz. polylectally. Theory - if it means explanation and prediction - depends on development and change, on ascertaining how structures come into being, and on a dynamic comparison of the variation patterns resulting from change. We must admit that it is development that explains states, not vice versa, and then either give up all hope of synchronic explanatory theories, or become developmentalists. This is the paradigmatic difference among frameworks today.

Edmund Gussmann: The so-called substance based approach is in fact also a formal approach, but formal in a different sense. In natural generative phonology certain theoretical restrictions and conditions are established on the basis of some external evidence. But then these restrictions are generalized and applied to other data for which no external evidence is offered or simply where the evidence is not available. This is, of course, perfectly legitimate, but it shows that Basbøll is not right in what he says in footnote 8 of his report. In fact, substance based phonologists proceed in exactly the same way as abstract phonologists, though their restrictions are largely phonetic. But this phonetic nature is, in fact, often avoided without any real justification. For example, the "true generalization condition" is exempt from applying in the case of different styles and tempos.

When professor Hooper claims that phonological rules should correspond to phonetic data in a predetermined way, then there is little for descriptive or practising phonologists to do, since we have here really some sort of discovery procedure.

The standard generative approach to the question of how much structure should be assigned to individual lexical items was autonomous by being divorced from rules of word formation. A number of problems could have been avoided, if the direction of morphological

processes had been taken into account. In some instances you can show that the rules of word formation have to take as their input the surface phonetic representation, in other cases the data argue just as strongly for abstract underlying representations as their input. There is a general non-existence of a theory of word formation. Here English seems to be a bad language to start with. In Slavic the very common expressive formations, such as augmentatives, diminutives, which are highly productive, are morphological processes which involve a number of phonological consequences. These should be studied in the first place, and rather than wondering whether 'serene' and 'serenity' are related. It is precisely in the interface of morphology, both inflectional and derivational, and phonology, that one should seek justification of phonological generalizations rather than in arbitrarily imposed restrictions of any sort.

Henning Andersen: Stephen Anderson's report seemed to me a very gracious concession of the total defeat of TG phonology. His remarks today seemed to contrive admission that it has not produced any results as a consequence of the monumental efforts made.

Basbøll's choice of leaving aside the vast amount of papers and monographs that contain important theoretical contributions under language-particular headings is regrettable. As to his limitation to descriptive linguistics, Bailey has taken care of that. But when Basbøll, in one of his footnotes, defines the substance based approaches as ones that go beyond the normal use of language, he must mean by that that they are interested in real data, meaning the use of phonology in speech, including speech errors, in verbal games, in poetics, by children, by aphasics, and so on.

In the same footnote, 'substance based' does not mean 'substance based' but rather 'speech based', - the traditional distinctions between language and speech, form and substance, etc. should be maintained also in discussions of these issues. I would like to ask Basbøll and Hooper to clarify what they mean by the distinction between formal and substantive, or if they understand them as being as vague as I do.

It is important to understand that language is something which is constantly changing, whose existence is in transmission from speaker to speaker, from generation to generation. Synchronic analysis is an artefact of the analyst. One must not identify

synchrony with the static, nor dynamism with diachrony: there can be dynamism in synchrony, and in diachrony you can talk about static facts, viz. the correspondences between two stages of a language.

In the transmission of language there are two logically distinct processes at work: deduction and abduction. Speakers know the grammar of the language and can produce deductively utterances which are correct. If you know the grammar, you can predict what sorts of utterances are going to be produced by that grammar. The other phase is the abductive one, by which speakers (children or adults) infer the grammar of the language from the speech they hear from speakers of the same dialect or from other dialects or even a foreign language. Logically, this is a process of hypothesis-making, about the content of the speech or about the grammar behind the speech. In this phase we cannot predict, but we can somehow understand the grammar. You cannot predict a grammar from the data, but you can form hypotheses about it. When we have constructed a grammar and understand that as a hypothesis, we can predict what sorts of innovation will be acceptable to speakers of that language, what sorts of verbal games will have which results, what kind of specific data would arise in aphasia - and we can test these hypotheses. On the other hand, given the speech data that learners of a language face when they acquire the language, we cannot predict the shape of the grammar they will produce. But we may be able to approach something like prediction if we understand that what they have to do in the process of arriving at a grammar is to make decisions, to form hypotheses. And if we understand that the data is susceptible to diverse analyses, contains ambiguities, we can capture these difficulties of analysis by formulating alternative hypotheses, and these hypotheses can then be subjected to empirical tests.

A proper theory of the ontology of language, which will be a proper theory both of synchrony and of diachrony, will enable us to both predict and to understand, will enable us to explain in both the senses that Bailey used, and hopefully future contributions of this kind will take in a wider scope of the field and see to what extent these various issues are faced by people working not specifically on descriptive linguistics but also on historical and pathological aspects of language, as well as the contributions made by people working in language-particular fields.

Joan Bybee Hooper: Gussmann says that if rules correspond to the phonetic substance in a predetermined way, then there is nothing for phonological theory to do. I think that is wrong. The formal theory may tell me what a rule is, given the phonetic data, it does not tell me how to figure out why there are these rules in particular rather than the other logically possible rules.

A clarification of the notion of substance: As an example we could consider the kind of criteria used in phonemic description; there are distributional criteria and then there is the criterion of phonetic similarity. Phonetic similarity would be a substantive criterion, while distribution would be considered formal. Another example: morphophonemics based on the properties of a morphological system would be a substantive approach, while morphophonemics treated as phonological would be a more formal approach.

Hans Basbøll: Synchronic linguistics seen as something absolutely static is a conception which I would not share.

Stephen Anderson: My view of the state of the SPE programme is that it proposed a particularly ambitious goal for constructing a logistic system that would reconstruct all of the content of sound structure. Certain fundamental inadequacies were clearly revealed in the comprehensiveness of the goals of that programme, as phonetic substance came to be taken more seriously into account. It seems to me that reactions to the perception of these failures have tended to throw out the baby with the bathwater and abandon the entire programme of SPE, and in particular its underlying rationalist assumptions, in an attempt to provide a rather radical sort of therapy for these problems. It seems to me that that is an overreaction; that one does indeed want to recognize that there are inadequacies in the attempt to reconstruct in such a logistic system all the content of phonology, but, nonetheless, one wants to preserve for that sort of system a central role in the development of phonology much as the sort of system in the Principia serves as a fundamental object of study within metamathematics.

Victoria Fromkin: The question is not: is the theory formal or substantive? but rather: is it a true theory of human language? I think that what Stephen Anderson has been trying to say is not that questions of articulation, etc., are not necessary for understanding certain aspects of language use, but that it is not necessarily the case that all aspects of language can be accounted

for by reference to these other aspects of language production and perception, etc. These questions of the philosophy of science are important because they have led us to look at different aspects which, hopefully, will eventually lead us to understand the nature of human language.

John J. Ohala: The issue of the psychological reality of phonological constructs has been raised during the discussion of this report and, in my opinion, has been made unnecessarily complex. I would like to simplify it with the following analogy, which is designed to appeal to the many academics in the audience. The problem of assessing the psychological reality of phonological constructs is very much like the problem the teacher faces in trying to verify that a student has mastered or knows the subject matter he has been exposed to in classes. How can this be done? Let us imagine three approaches: the teacher that takes the 'formalist' approach will just speculate on what it is possible for a student to know and will assume that that is what all students know. The teacher who would have most in common with those phonologists who have here been characterized as accepting 'substantive' evidence, would rely on additional 'external' evidence of a student's knowledge, e.g., what books he had in his library, whether he nodded sagely during the teacher's lectures, laughed at his jokes, etc. The teacher who would take the experimental approach would demand of all students some behavioral evidence that they had mastered the subject matter, e.g., performance on a written or oral test, an original paper or thesis, etc. Naturally this performance should not be attributable to anything other than the student's full mastery of the subject, e.g., cheating or random selections of answers to 'true/false' questions. I leave it to all those academics in the audience to decide which approach they would use. I would hope that whatever decision they make, however, that this would influence their practice in phonology, too.

The point is that different types of evidence in phonology vary considerably in their ability to unambiguously tell us what is in the speaker's head. Most of the evidence characterized as 'substantive' in this discussion, e.g., speech errors, sound change, is quite ambiguous in this regard. Only evidence from tests (experiments) can be minimally ambiguous. This is not to

say that there cannot be a bad test. But the proper response to a bad test - both in academia and in phonology - is an improved test. Teachers expend considerable time and imaginative effort refining the tests they use to assess the psychological reality of students' knowledge. Why shouldn't similar effort bear fruit in phonology?

Natalie Waterson: I should like to draw attention to another theoretical approach: to Prosodic Phonology initiated by J.R. Firth in England. Very briefly: most phonological theories have phonemic segments as the basic units of description, whether explicit or implicit, yet there is general recognition by those who study speech perception that the phoneme has yielded little in the way of insights to our understanding of how speech is perceived and interpreted, and it is becoming plain that it is not the right unit for such studies. In Prosodic Phonology the unit of description is the word, phrase, or sentence, and features which synthesize the word, etc., into a whole as well as those that divide it up are taken into account, i.e. syntagmatic and paradigmatic relations.

The phonological system of a language is thus described in terms of different word, etc., structures and not in terms of a system of phonemic segments. No exposition of the theory is available but there is plenty of illustrative material in theses and papers produced in the Dept. of Phonetics and Linguistics, at SOAS, University of London. Most of the material is about Oriental and African languages and the only English material are my papers on child phonology.

It is interesting to see the influence of Prosodic Phonology on developing theories, for instance on Joan Bybee Hooper's approach, and autosegmental phonology.

Richard Coates: The SPE type of phonology, represented here by professor Anderson, has tended to specify a kind of codified norm, whereas professor Hooper's system specifies the linguistic rules which would characterize usage as being the starting point of changes. I think that together they comprise the native speaker's system, both a kernel, or norm, available to him, and a system of partly specified potential directions of the changes. Thus, the output of morphology would not be absolutely rigidly defined, and we may imagine a speaker who makes very few morpho-

logical connections between surface forms not connected by phonological rules, on the one hand, and on the other a speaker who fluently manipulates a morphological and phonological system (à la James Foley's native speaker).

Wiktor Jassem: Fifteen years ago, or more, three points were made about generative phonology: observational adequacy, descriptive adequacy, and explanatory adequacy. Now, in the old days so little observation was done that it is difficult to say whether it was adequate or not; descriptive adequacy described rather what was going on in the minds of the theorists; explanatory adequacy, for which the criterion was simplicity, led to rules which in structural phonology could be expressed by three or four symbols but which in TG took complete pages so full of things that you could not see the wood for the trees. My point is: I suppose that revolution in phonology did not start twenty or seventeen years ago with Chomsky, - revolution in phonology, according to what I have heard today and read in the Proceedings, is starting now!

Royal Skousen: Each approach to phonology proposes a method of analysis. In some sense they are all formal in that they look at the data and attempt to derive a description from the data, but I would prefer to call that a method of induction or learning. I would like to suggest that, in addition to these formal considerations or these principles of learning, there is a need also for an empirical interpretation of the description: What does my description actually predict about language usage, about language intuition? - Furthermore, we need first to explicitly determine how we get our description from the data, and secondly, to answer the question of what would convince us that our description is right or wrong, because in the absence of such arguments we do not really have a theory at all.

William Haas: There is another kind of opposition that has to be reconciled, namely the opposition between empirical and speculative. More than twenty years ago, Martinet published his "Phonology as functional phonetics". And that was a kind of reconciliation: phonology was to present criteria for relevance, criteria of selection, to apply to the mass of unorganized phonetic data. Now we seem to have had some fifteen years of something different: phonology as speculative phonetics, and we

are now not so much imposing criteria of relevance on phonetic research as asking the phonetician to provide us with criteria to decide amongst different formal systems of phonology. Amongst these criteria will be the old functional phonology which is now, as it were, part of the surface data.

Kenneth L. Pike: It is not possible to separate phonology from grammar, from lexicon, from meaning. We must have a tri-hierarchical structure: phonology, grammar, and meaning. But in each of the hierarchies there are thresholds. - No mathematical system of any complexity can be treated as consistent by looking at the data inside itself. Something external must be used. That which I use from outside the formal system, to make it relevant, is meaning and behavioral impact.

Hans Basbøll: I want to stress once more that if my report is to be read as a status report on phonology, it should be read in connection with the contributions to the symposia.

Stephen Anderson: Perhaps we can all agree that the fundamental problem for phonologists is the exploration of what can constitute the sound pattern of a language. Ultimately we all have to make our own choice about what is the most productive way to go about this investigation, and I think it is unlikely that there are determinate answers to the sorts of opposition questions that have been posed.

## THE RELATIONS BETWEEN AREA FUNCTIONS AND THE ACOUSTICAL SIGNAL

Gunnar Fant, Department of Speech Communication, Royal Institute of Technology, S-10044 Stockholm, Sweden

Chairpersons: Wiktor Jassem and Kenneth N. Stevens

Introduction

The topic of this paper is to discuss how configurations, shapes, and detailed outlines of the vocal tract cavity system influence the acoustic signal and the reverse, how to predict vocal tract resonator dimensions from speech wave data. As far as the direct transform is concerned, this is a re-visit to my old field of acoustic theory of speech production.

What progress have we had in vocal tract modeling and associated acoustic theory of speech production during the last 20 years? My impression is that the large activity emanating from groups engaged in speech production theory and in signal processing has not been paralleled by a corresponding effort at the articulatory phonetics end. Very little original data on area functions have accumulated. The Fant (1960) Russian vowels have almost been overexploited. Our consonant models are still rather primitive and we lack reliable data on details of the vocal tract as well as of essential differences between males and females and of the development of the vocal tract with age.

The slow pace in articulatory studies is of course related to the hesitance in exposing subjects to X-ray radiation. Much hope was directed to the transformational mathematics for deriving area functions from speech wave data. These techniques have as yet failed to provide us with a new reference material. The so-called inverse transform generates "pseudo-area functions" that can be translated back to high quality synthetic speech but which remain fictional in the sense that they do not necessarily resemble natural area functions. Their validity is restricted to non-nasal, non-constricted articulations and even so, they at the best retain some major aspects of the area function shape rather than its exact dimensions. However, some improvements could be made if more representative acoustic models than LPC analysis are considered.

Once a vocal tract model has been set up it can be used, not only for studying articulation-to-speech wave transformations, but also for a reverse mapping of articulations and area functions to fit specific speech wave data. These analysis-by-synthesis re-



mapping techniques, as well as perturbation theory for the study of the consequences of incremental changes in area functions or of the inverse process, are useful for gaining insight in the functional aspect of a model. However, without access to fresh articulatory data the investigator easily gets preoccupied with his basic model and the constraints he has chosen.

The slow advance we have had in developing high quality synthesis from articulatory models is in part related to our lack of reliable physiological data, especially with respect to consonants, in part to the difficulty involved in modeling all relevant factors in the acoustic production process. The most successful attempt to construct a complete system is that of Flanagan et al. (1975) at Bell Laboratories. A variety of studies at KTH in Stockholm and at other places have contributed to our insight in special aspects of the production process such as the influence of cavity wall impedance, glottal and subglottal impedance, nasal cavity system, source filter interaction, and formant damping.

From area function to the acoustic signal

The acoustic signal or, in other words, the speech wave is the product of a source and a filtering process. The most common approach is to disregard the source and relate a vocal tract area function to a corresponding formant pattern only, i.e. a set of formant frequencies  $F_1, F_2, F_3, F_4$ , etc. This correspondence is illustrated by Fig. 1. I shall not go into the mathematics of the wave equations and the equivalent circuit theory. Instead I will attempt to develop a perspective around some basic models and current problems.

To derive an area function from X-ray data on vocal tract dimensions is by no means a straightforward procedure, see Fant (1960; 1965) and Lindblom and Sundberg (1969).

The estimation of cross-sectional shapes and dimensions in planes perpendicular to the central pathway of propagation through the vocal tract has to rely on crude conventions and involves uncertainties, e.g. with respect to variations with articulation and for different types of subjects. The lack of basic data is especially apparent for female and child speech and for consonants, e.g. laterals and nasals. In spite of the accessibility of the speech wave to quantitative analysis there is a similar lack of reference data concerning the acoustic correlates. Most studies have been concerned with male speech and vowels.

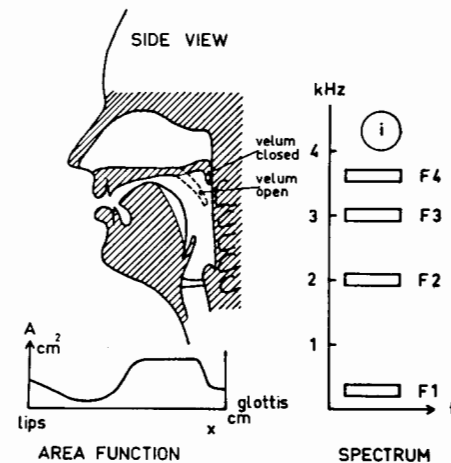


Figure 1. Principle illustration of vocal tract sagittal view with area function and corresponding resonance frequency pattern.

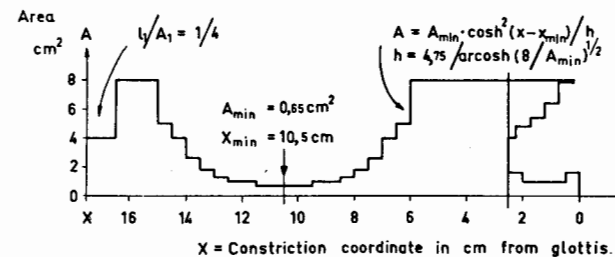


Figure 2. Three-parameter vocal tract model (Fant, 1960).



A specification of an area function as a more or less continuous graph of cross-sectional area from the glottis to the lips allows detailed calculations of the acoustic response but is not practical for systematic descriptions. A data reduction in terms of parametric models brings out the acoustically relevant aspects. The three-parameter models of Stevens and House (1955) and Fant (1960) differ somewhat in the details but have the same set of parameters, the place of minimum cross-sectional area of the tongue section, the area at this coordinate, and the length over area ratio  $l_0/A_0$  of the lip section.

My model is shown in Fig. 2. The shunting sinus piriformis cavity around the outlet of the larynx tube was a constant feature in my model. A weakness is that it is not reduced in volume for back vowels which does not allow  $F_1$  to reach a sufficiently high value for [a]. Fig. 3 shows the variation of the F-pattern with the place of tongue constriction. This is a well established graph which retains basic patterns such as the rise of  $F_2$  with advance of the tongue constriction from back to front up to an optimal place at a midpalatal location after which  $F_2$  drops again. A limitation of the parameter range to a region bounded by [a], [u], and [i] as proposed in several articulatory models, e.g. Lindblom and Sundberg (1969), would exclude the standard Swedish pronunciation of the vowel [ɛ] which, contrary to traditional classifications, has a constriction somewhat anterior to that of [i] (Fant, 1973).

The constriction coordinate is an acoustically more relevant classifier than the "highest point of the tongue" of classical phonetics. Most stressed vowels have a definite "place of articulation" as evidenced by a region of minimum cross-sectional area which we may exemplify by [i], [u], [o], [a] ending with a variant of [ɛ] with major narrowing just above the glottis (Fant, 1960). On the other hand, it may be argued that the traditional classification in terms of tongue locations and related parameters belongs to a production stage one step higher up than area functions and could be directly related to formant patterns.

The [a] and [i] vowels are polar opposites, the [i] vowel requiring a wide pharynx and narrowed mouth, whilst the opposite is true of [a] type vowels. A production of a vowel [u] requires a double resonator configuration with a narrow lip opening to ensure

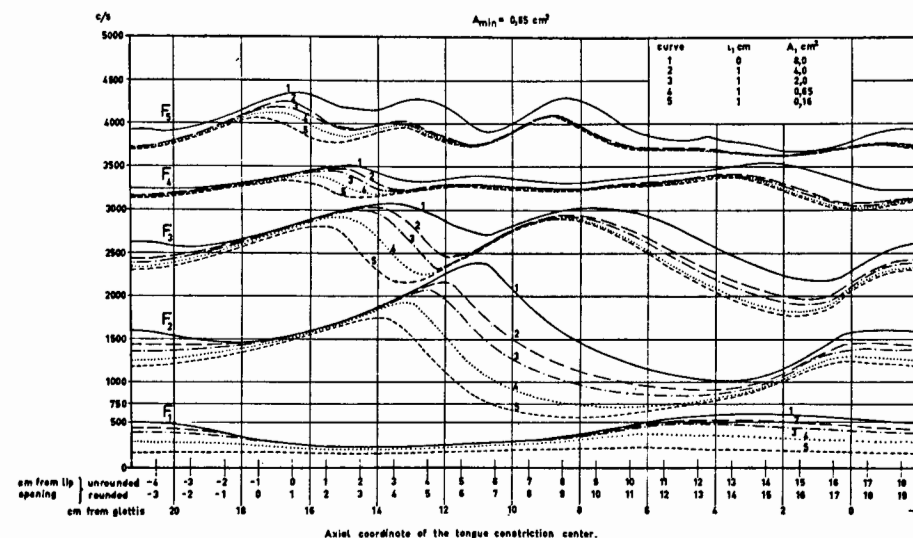


Figure 3. F-pattern variation with constriction coordinate  $x_c$  at different sets of lip parameter  $l_1/A_1$  at constant constriction area  $A_{min}$ . The constriction coordinate is zero at the glottis.

a low  $F_1$  and a narrow constriction between the two major cavities as a correlate of a low  $F_2$ . These shape aspects are brought out in the stylized area functions of Fig. 4. A basic issue in acoustic phonetics is that it is not possible to produce these vowels without retaining the major shape aspects of the area functions. To this extent area functions are predictable from the acoustic signal as will be discussed in greater detail in a later section. Peter Ladefoged would back me up here with his competence of transforming phonetic qualities to equivalent resonator configurations.

Another basic issue is that the vocal tract filtering is determined by the location of formants only and that the spectrum envelope between peaks cannot contain any other irregularities than those originating from the source function. Minor irregularities in the outline of the area function may have some influence on formant locations but will not give rise to irregularities in the spectrum envelope. This is not evident without an insight in the mathematical constraints imposed by acoustic theory. It is related to the one-dimensional wave propagation, wavelengths generally being short compared to vocal tract cross dimensions. Systematic perturbations of vocal tract area functions will be discussed in a later section.

Highly simplified area functions of fricatives (or corresponding stops) and their filtering functions are shown in Fig. 5. As discussed by Fant (1960), the "compact" sibilant [ʃ] or the stop [k] has a definite cavity in front of the major constrictions which accounts for a central dominance of the spectrum, usually a single formant, if the cavity is abruptly terminated by the constriction. The [s] or [t] has a narrow channel of a few centimeters length behind the source which may combine with a small front cavity to produce resonances above 4000 Hz which build up a high-pass filtering. The [f] or [p] has no significant resonance in its closed state.

In general, the cavities behind the source do not influence the spectrum much, provided that the consonantal constriction is effective. Resonances of the back cavities may appear if the constriction tapers off gradually as in palatals or if a palatal tongue articulation builds up a supporting constriction behind the lips. Back cavity resonances combine with and are cancelled by spectral zeroes at complete closure but move away from their

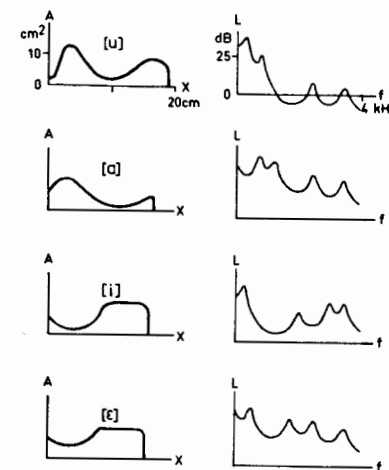


Figure 4. Stylized area functions and corresponding spectrum envelopes of [u] [a] [i] and [ε]. The constriction coordinate is zero at the lips.

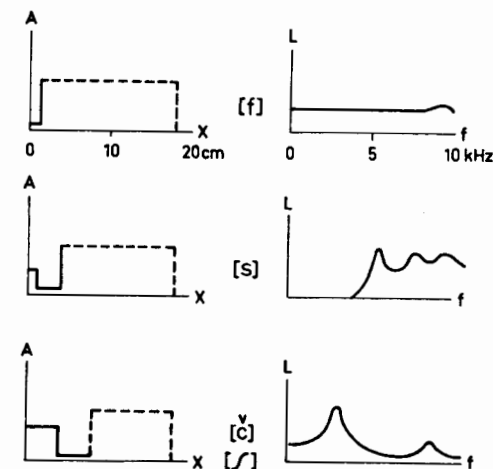


Figure 5. Stylized area functions and corresponding spectra of three basic consonant categories. The constriction coordinate is zero at the lips.

zero mates during release and are then more or less free to appear. In Fig. 6 we can study measured and calculated spectra of [k] and a palatalized [p'] (Fant, 1960). The labial burst spectrum contains peaks at around 2-3 kHz but has a free spectral minimum at 1400 Hz. In contrast, the [k] spectrum has a single formant peak around 1400 Hz. It is interesting to note that the calculations from the area function data back up the measured spectra. We need more studies of this type.

#### Vocal tract boundary constraints and dynamics

The simplified static models relating a single area function without parallel branches to a set of formant frequencies have obvious limitations. On a higher level of ambition we must include proper boundary conditions such as radiation load and a finite coupling to the subglottal and nasal systems. In order to predict formant bandwidths we must consider the energy loss during an oscillatory cycle of a formant associated with "loss elements" on the surface of the vocal tract resonator system and other dissipative elements (Fant and Pauli, 1975). Source functions must be defined with respect to place of insertion in the vocal tract, their spectrum or waveform, and the degree of coupling to other parts of the system (Stevens, 1971). In addition, these properties are highly time variable within a voice fundamental period (Fant, 1979) and within intervals of transition from various states of the glottis or of other terminations of the vocal tract. Rapid opening and closing gestures pose specific problems in relating area functions to acoustic data. In a proper analysis of connected speech we need two sets of acoustic variables: the continuous variations of the F-pattern as a correlate of the continuous movements of the articulators and the often abruptly varying patterns of spectral energy distributions associated with discrete events of production.

The acoustic production model of Fig. 7 may serve as a starting point for a brief discussion of these problems. First of all, we should note an important element in converting area functions to a filter function. The walls of the vocal tract are not rigid. They may expand during a voiced occlusion as represented by the element  $C_w$  in the equivalent circuit of a small slice of the area function, Fig. 8, and they have a finite mass  $L_w$  which adds to the tuning of vocal resonances and which dominates the impedance of

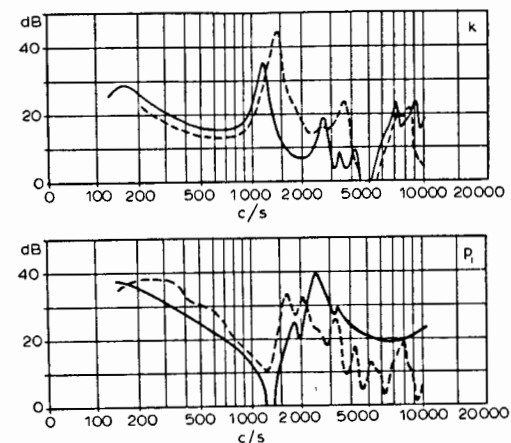


Figure 6. Calculated (solid line) and measured (broken line) stop release spectra of a velar [k] and a palatalized [p']. The minimum in [p'] at 1400 Hz is a free zero in the sub-lip impedance whilst the main formant of [k] is a mouth cavity formant. After Fant (1960).

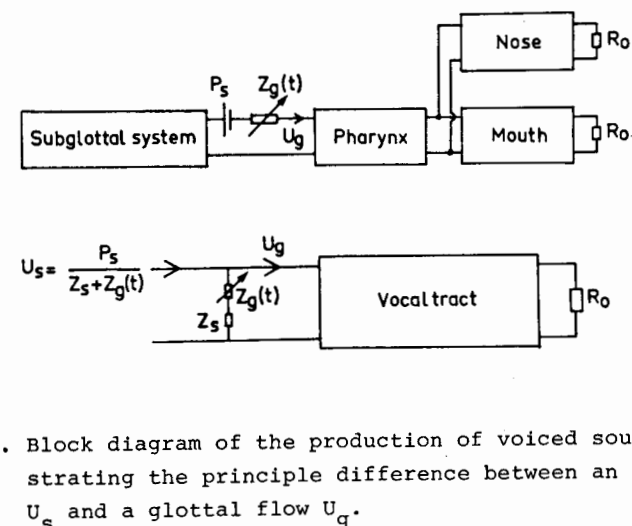


Figure 7. Block diagram of the production of voiced sounds illustrating the principle difference between an ideal source  $U_s$  and a glottal flow  $U_g$ .

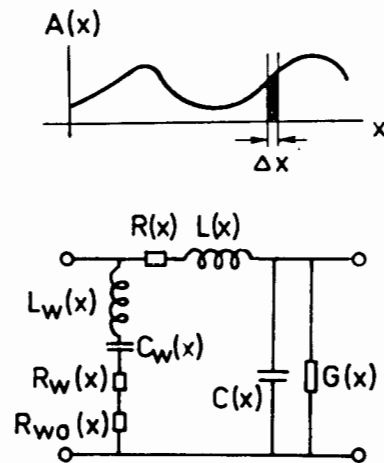


Figure 8. Lumped constant approximation of a small slice of the area function.

the shunting branch at frequencies above 40 Hz. A small fraction of sound is radiated externally from the outside of the head through  $R_{wo}$ . It is negligible except as a constituent of the voice bar of a voiced occlusion.

Disregarding the cavity wall mass element  $L_w$ , calculations would provide  $F_1 = 0$  for an area function starting and ending with complete closure. The finite  $F_1$  of around 150-250 Hz found in the spectrogram of the voiced occlusion is determined by the resonance of the entire air volume compliance in the tract with the total lumped cavity wall mass shunt. This resonance can easily be measured acoustically (Fant et al., 1976) and amounts to  $F_{1w} = 190$  Hz with a bandwidth of  $B_{1w} = 75$  Hz, typically for a male voice, and around 20% higher for females. The wall mass element  $L_w$  is thus an important constituent in calculating  $F_1$  from the area function. The procedure is to start out with a derivation of an ideal  $F_{1i}$  without mass shunt and add a correction factor

$$F_1 = F_{1i} (1 + F_{1w}^2 / F_{1i}^2)^{1/2} \quad (1)$$

The distribution of the wall impedance along the vocal tract and its dependence on particular articulations are not known. The experiments of Fant et al. (1976) suggest that regions around the larynx and the lips are especially important. Experiments by

Ishizaka et al. (1975) provide data of the same order of magnitude but have not revealed conclusive distribution patterns.

The resistive component  $R_w$  in the cavity wall branch determines a major part of the bandwidth  $B_1$  of low  $F_1$  formants. The resistive part of the radiation load which is proportional to frequency squared is the essential bandwidth determinant of resonances above 1000 Hz originating from an open front resonator. Internal surface losses from friction and heat conduction enter through the elements  $R$  and  $G$  in Fig. 8. They are proportional to the half power of frequency and to the inverse of the cross-sectional area. A detailed analysis of formant bandwidths and their origin appears in Fant (1972), Fant and Pauli (1975), and Wakita and Fant (1978).

The time variable glottal impedance accounts for variations of formant frequencies and bandwidths within a voice fundamental period (Flanagan, 1965). A more detailed analysis of glottal damping requires a reconsideration of the process of voice generation (Fant, 1979) and adoption of perceptual criteria for deriving equivalent mean values (Fant and Liljencrants, 1979). The main excitation of the vocal tract occurs at the instant of interruption of glottal flow by glottal closure. At this instance, damped oscillations are evoked and subjected to the damping from supra-glottal loss elements.

When the glottis opens for the next flow pulse the vocal tract becomes loaded by the time variable glottal plus subglottal impedance. Providing a resonance mode is much dependent on the part of the area function immediately above the glottis, the glottal damping becomes severe. This is especially apparent if the lower pharynx is narrowed thus facilitating an impedance match between the cavity system and the glottal resistance. A complete extinction of the formant oscillation in the glottal open interval may result. This is typical of  $F_1$  of the vowel [a] produced at low or moderate voice effort by a male subject.

In general most of the energy excited during a voice fundamental period is lost during the timespan of the following period. Since glottal resistance decreases with lowered transglottal pressure the damping effect is especially apparent at weak voice levels. The mean glottal bandwidth in normal voice production is of the order of 0-100 Hz with 20 Hz as a typical value for male medium intensity phonation.

It is apparent that any model of voice production which adopts the actual flow through the glottis as the primary source will create problems. With this convention, which happens to apply to inverse filtering techniques, the source attains components of formant oscillations and becomes dependent of the vocal tract area function (Mrayati and Gu erin, 1976). Their approach is intended to define a proper source for a formant synthesizer.

A different approach more suited for production models is to incorporate the combined glottal and subglottal impedance as a termination paralleling the input end of the tract and to define the source as the flow through the glottis which would have occurred with the input to the vocal tract short circuited. This representation adopted by Fant (1960) preserves a realistic definition of the vocal tract transfer function but fails to take into account source modifications due to aerodynamic losses in supraglottal constrictions. In the transition from a vowel to a voiced consonant there is generally some loss of transglottal pressure which reduces the excitation strength of the voice source.

The interplay of glottal and supraglottal sources associated with articulatory narrowing and release becomes an important part of a dynamically oriented theory of predicting acoustic signals from area functions (Stevens, 1971).

What about the subglottal system? How does it influence speech? In normal voice production the influence appears to be small. As long as the glottal opening is small and the flow velocity high, the glottis impedance becomes high compared to the subglottal impedance. Unless there is a constant leakage bypassing the vibrating part of the glottis, the subglottal system should have a minor influence only.

This reasoning is concerned with the modification of the supraglottal formants only. At the instance of flow interruption when the glottis closes there is a simultaneous excitation of resonances in the trachea and other parts of the subglottal system. Potential frequencies are 600, 1250, and 2150 Hz for a male voice (Fant et al., 1972). The transmission losses associated with the penetration of these components through the walls of the trachea and the chest to externally radiated sound appear to be sufficiently high to rule out any significance, but this remains to be proved.

As shown by Fant et al. (1972), subglottal formants may occasionally be seen in spectra from aspirated sound segments, e.g. in the release phase of unvoiced stops. "F1-cutback" in the first part of the voiced interval after release, which appears as a relative delay in onset of F1 compared to F2 and higher formants, may be explained as an instance of excessive F1 damping through an incompletely closing glottis. The upper formants are less dependent on the glottal termination and thus less affected. This relative weakening of F1 is a filtering effect, whilst the relative weakness of F1 in a preceding unvoiced, aspirated segment is also a matter of low source energy in the F1 region. The F1 intensity reduction is also seen in the terminating periods of a vowel before the occlusion of an unvoiced stop (pre-occlusion aspiration).

Nasalization and aspiration have similar effects on F1. In nasalized sounds the F1 intensity is typically reduced by a spectral zero (Fant, 1960; Fujimura and Lindqvist, 1971). The nasal model of Fant (1960) produces too high values of the lowest nasal pole. The possible occurrence of several low frequency pole-zero pairs is made plausible by the study of Lindqvist and Sundberg (1972). More anatomical and acoustic data are needed.

In connection with the voice source studies of Fant (1979) it has been noted that the spectral maximum often seen below  $F_1$  in vowels is a voice source characteristic, which becomes especially enhanced in contrast to a weak F1 in nasalized or aspirated, voiced segments. This is especially apparent in a time domain study. Another way of expressing this finding is to say that nasal sounds retain more source characteristics than non-nasal sounds.

If an area function is subjected to a substantial change in a very short time, one may expect some deviations from the linear stationary behavior. Point-by-point calculations of resonance frequencies are still valid but additional bandwidth terms enter which may be positive or negative. A rapid opening of a constriction is accordingly associated with a negative bandwidth component and a rapid closure with a positive bandwidth component. The analysis is simple. Consider a flow  $U(t)$  through an acoustic inductance  $L(t) = \rho l/A(t)$ . The pressure drop is:

$$P(t) = \frac{d}{dt} [L(t)U(t)] = L'U + LU' \quad (2)$$

$L' = dL/dt$  apparently has the dimension of a resistance  $R_d$

$$R_d = \frac{dL}{dt} = \frac{-A'(t)\rho l}{A^2(t)} \quad (3)$$

In a single resonator system the bandwidth component associated with a resistance  $R$  in series with an inductance  $L$  is simply  $R/2\pi L$ .

Accordingly, the bandwidth associated with  $R_d$  is

$$B_d = \frac{-A'(t)}{2\pi A(t)} \quad (4)$$

which implies a bandwidth component of opposite sign to that of the rate of change of the area. Fig. 9 illustrates the temporal course of the bandwidth when a resonator of volume  $100 \text{ cm}^3$  is coupled to a neck of length  $4 \text{ cm}$  and a cross-sectional area  $A(t)$  varying exponentially from closure to complete opening of  $2 \text{ cm}^2$  with a time constant of  $10$  milliseconds.

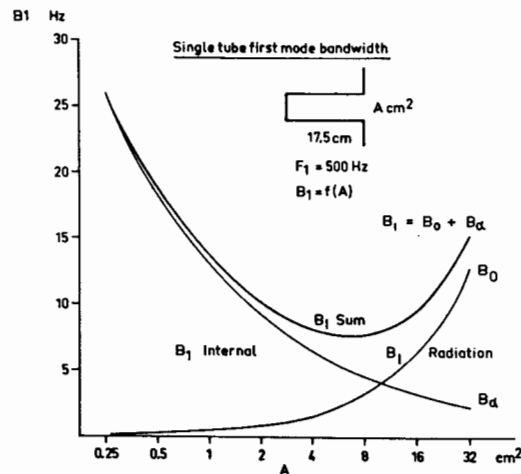


Figure 9. Resonator outlet area  $A$ , resonance frequency  $F$ , and total bandwidth  $B$  as a function of time during an exponential release with a time constant of  $10$  milliseconds.  $B_d$  is the negative dynamic component of the bandwidth.

The time varying negative bandwidth overrides the frictional bandwidth components up to  $8$  milliseconds after release which could tend to increase the amplitude of the oscillation during that period. However - in the speech case there enter additional positive bandwidth components related to flow dependent resistance and to cavity wall losses and possibly also glottal losses which tend to reduce the importance of the negative terms. In a detailed analysis of the glottis resistance the dynamics calls for some decrease of glottal resistance in the rising branch of the glottal pulse and an increase in the falling branch, as noted by Guérin et al. (1975). Except for the analysis above, a proper evaluation of the practical significance has to my knowledge not been performed. The most detailed thesis on the theoretical aspects is that of Jospa (1975). I feel that dynamic effects are of academic rather than practical significance. Of greater importance is probably the mere fact that a rapid transition of a formant creates a special perceptual "chirp" effect.

Perturbation theory and vocal tract scaling

Perturbation theory describes how each resonance frequency,  $F_1, F_2, F_3$ , etc., varies with an incremental change of the area function  $A(x)$  at a coordinate  $x$  and allows for a linear summation of shifts from perturbations over the entire area function. The relative frequency shift  $\delta F/F$  caused by a perturbation  $\delta A(x)/A(x)$  is referred to as a "sensitivity function". We may also define a perturbation  $\delta \Delta x/\Delta x$  of the minimal length unit  $\Delta x$  of the area function which will produce local expansions and contractions of the resonator system. It has been shown by Fant (1975b), Fant and Pauli (1975) that the sensitivity function for area perturbations of any  $A(x)$  is equal to the distribution with respect to  $x$  of the difference  $E_{kx} - E_{px}$  between the kinetic energy  $E_{kx} = \frac{1}{2}L(x)U^2(x)$  and the potential energy  $E_{px} = \frac{1}{2}C(x)P^2(x)$  normalized by the totally stored energy in the system.

Fig. 10 from Schroeder (1967) illustrates perturbations of a single tube resonator by changes in the area function derived from sinusoidal functions. These have been chosen to influence  $F_1$  only (a), none of the formants (b), and  $F_2$  only (c). The middle case is of special interest. There exists an infinite number of small perturbations applied symmetrically with respect to the midpoint of the single tube, which will have almost no influence

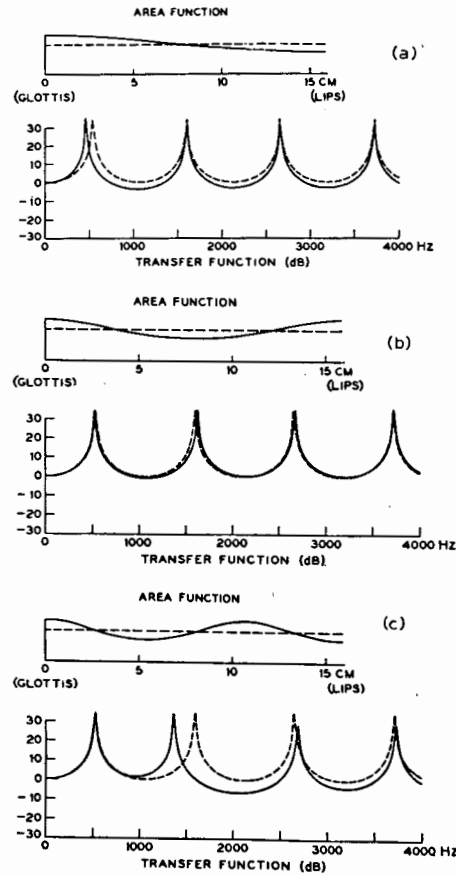


Figure 10. Perturbations of the single tube area function affecting  $F_1$  only (a), almost no influence (b), and  $F_2$  only (c) (after Schroeder, 1967).

on the formant pattern. In the general case of an arbitrary area function the rule of symmetry is upset (Heinz, 1967) but there still exists a tendency of compensatory interaction between front and back parts (Öhman and Zetterlund, 1975).

Sensitivity function for area perturbations of my six Russian vowels are shown in Fig. 11. This chart is useful as a reference for general use. Given the relative amount of area change, the corresponding relative frequency shift  $\delta F_n / F_n$  is proportional to the product of  $\frac{\delta A(x)}{A(x)}$  and the amplitude of the sensitivity function,  $E_{kx} - E_{px}$ . As an example we may note that  $F_1$  of the vowel [u] rises with increasing area at the lips, i.e. decreases with increasing degree of narrowing and that narrowing the tongue constriction of [u] causes  $F_2$  to fall and  $F_3$  to rise. A narrowing of the outlet of the larynx tube will apparently have the effect of tuning  $F_4$  to a lower frequency.

With the area function sampled at intervals of  $\Delta x$ , e.g.  $\Delta x = 0.5$  centimeter for practical use, we may ask what happens if we increase  $\Delta x$  at the coordinate  $x$  by the amount  $\delta \Delta x$ . The local expansion thus introduced causes a frequency shift  $\delta F_n / F_n$ , which is proportional to  $-\delta(x) / (1 + \delta(x))$  and to  $(E_{kx} + E_{px})$  of resonance  $n$ .

The distribution of  $(E_{kx} + E_{px})$  is uniform for a single tube resonator. The effect of a length increase is obviously the same irrespective of where along the  $x$ -axis the tube is lengthened. An overall increase of the length by, say  $\delta(x) = 0.2$ , causes a shift of all resonance by a factor  $-0.2 / (1 + 0.2) = -0.17$ . The same calculation performed directly from the resonance formula  $(2n-1)c / 4l_t$ , where  $l_t$  is the total length and  $c = 35300$  cm/s is the velocity of sound, would provide the same answer, i.e. a frequency ratio of  $1 / (1 + 0.2) = 0.83$ .

The distribution  $E_{kx} + E_{px}$  along the vocal tract is also a measure of the relative dependence of the particular resonance mode on various parts of the area function. This is the best definition we have of "formant-cavity" affiliations. From Fig. 12 we may thus conclude that most of the energy of the second formant of [i] is stored in the pharynx, whilst the third formant of [i] "belongs to" the front part of the system.  $F_3$  of the back vowels [u] [o] and [a] are associated with a central part of the tract, and  $F_4$  of all vowels has a substantial peak of energy located in

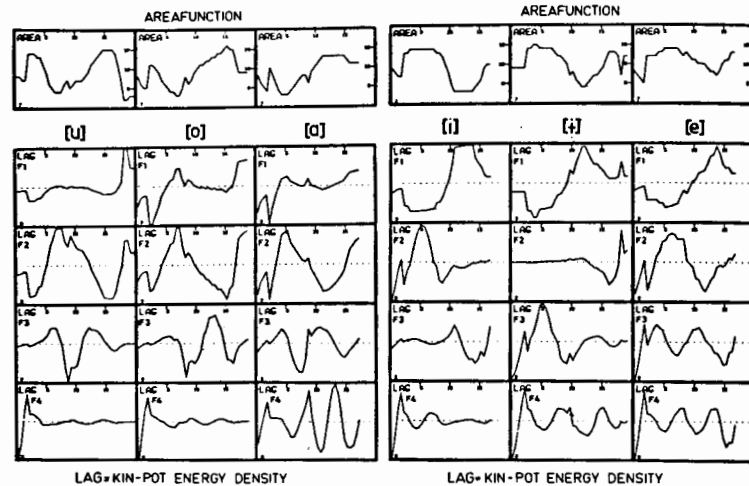


Figure 11. Sensitivity functions for area perturbations of the six Russian vowels (Fant, 1960). From Fant (1975b). The constriction coordinate is zero at the glottis.

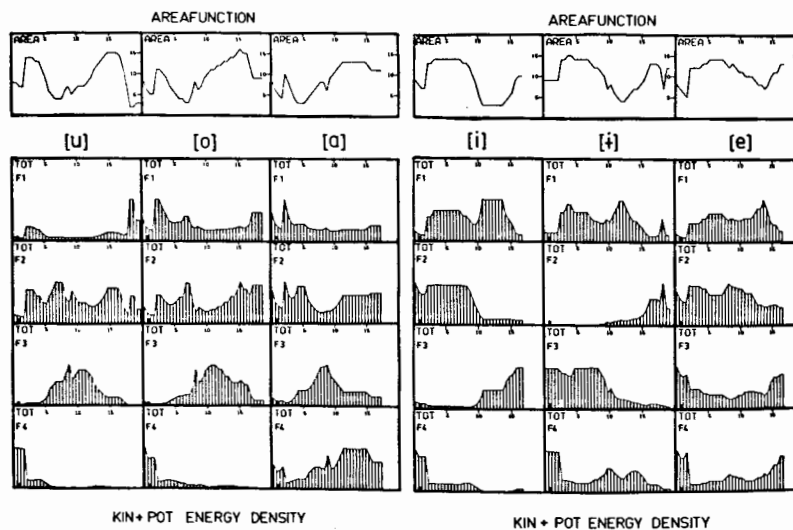


Figure 12. Sensitivity functions for length perturbations of the six Russian vowels (Fant, 1960). From Fant (1975b). The constriction coordinate is zero at the glottis.

the larynx tube. Expanding the length of the pharynx will have a large effect on  $F_2$  of [i] and a small effect only on  $F_3$  and vice versa for a length expansion of the mouth cavity. This analysis would apply to the relatively short pharynx of females compared to males.

If a perturbation of the entire area function is expressed as a function of as many parameters as there are formants, it is possible to calculate the change in area function from one F-pattern to another (Fant and Pauli, 1975). This indirect technique has been used by Mrayati et al. (1976) for deriving plausible area functions for French vowels on the basis of their deviation from my reference Russian vowels. This procedure must be administered in steps of incremental size with a recalculation of the sensitivity function after each major step. It may involve length as well as area perturbations.

In practice, when aiming at direct transforms only, it may be easier to resort to a direct calculation of the response of the perturbed area functions than to derive it from the energy distributions. The perturbation formulas and especially their energy based derivations are more useful for principal problems of vocal tract scaling or for gaining an approximate answer to a problem without consulting a computer program.

The area functions of male and female articulations of the Swedish vowels [i] and [u] and corresponding computed resonance mode pattern in Fig. 13 may serve to illustrate some findings and problems. The data are derived from tomographic studies in Stockholm many years ago in connection with the study of Fant (1965; 1966) and were published in Fant (1975a; 1976). It is seen that in spite of the larger average spacing of formants in the female F-pattern related to the shorter overall vocal tract length, the female  $F_1$  and  $F_2$  of [u] and the  $F_3$  of [i] are close to those of the male. This is an average trend earlier reported by Fant (1975a), see Fig. 14. Differences in perceptually important formants may thus be minimized by compensations in terms of place of articulation and in the extent of the area function narrowing. Such compensations are not possible for all formants and cannot be achieved in more open articulations. The great difference in  $F_2$  of [i] is in part conditioned by the relatively short female pharynx but can in part be ascribed to the retracted place of



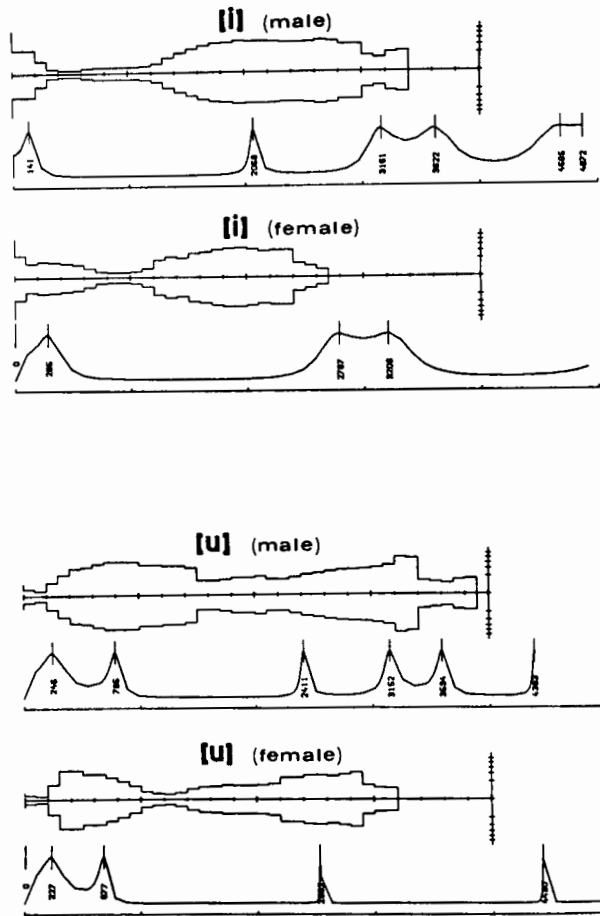


Figure 13. Male and female vocal tracts (equivalent tube representation) and corresponding F-patterns from the tomographic studies of Fant (1965).

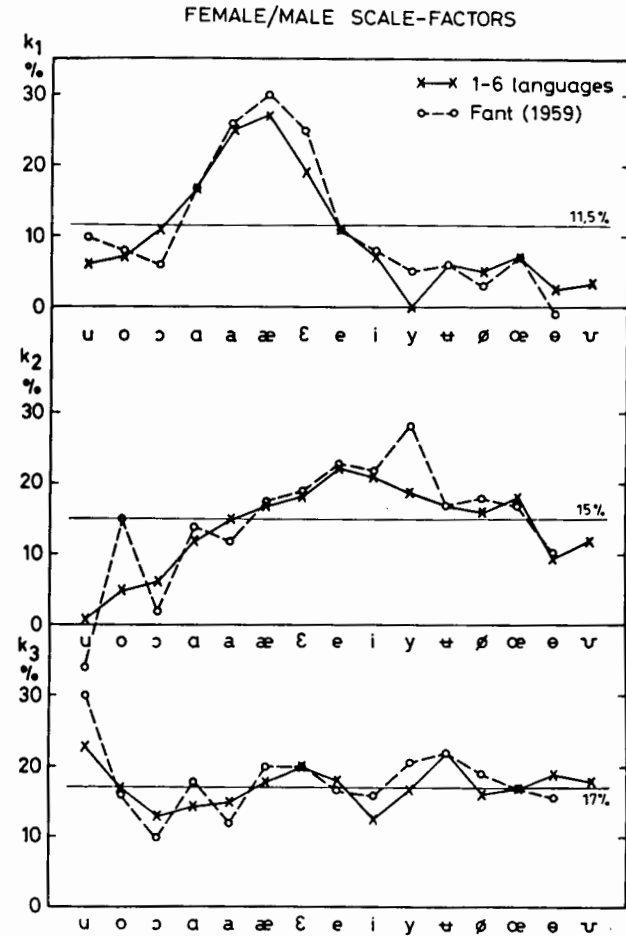


Figure 14. Female/male scale factor variation with vowel and the particular formant (Fant, 1975a).

articulation. It is also disputable whether this particular female articulation serves to ensure an acceptable [i] or whether there is a dialectal trend towards [ɪ]. Also, it is to be noted that X-ray tomography may impede the naturalness of articulations because of the abnormal head position required.

Much remains to be studied concerning how the vocal tract area functions of males, females, and children are scaled in actual speech and what kind of compensation occurs for minimizing perceptual differences or maybe the reverse, to mark contrasts between age and sex groups.

The lack of reference data on area functions is severe and the attempts to overcome this lack by means of area function scaling performed by Nordström (1975) were not conclusive except to support the general issue that the vowel and formant specific female-male differences, documented by Fant (1975a), Fig. 14, do not always come out as a result of the particular scaling assumed. The agreement was good for  $F_3$  and fair for  $F_2$  and rather bad for  $F_1$ . The predictability of  $F_3$  is expected in view of the high dependency of  $F_3$  on length dimensions.

A weakness in the Nordström study is that his [æ] and [ɛ] vowel area functions were interpolated from the Russian [ɑ] and [e] vowel and accordingly attain a centralized quality not representative of the [a] and [æ] category vowels which normally display a very large female-to-male  $F_1$  ratio, see Fig. 14.

It is interesting to note that the non-uniform differences between females and males are paralleled by similar patterns comparing tenor and bass male singers. These vowel and formant specific trends are not only the automatic consequence of different anatomical scalings but also reveal compensations according to criteria that are not very well understood yet. A promising project on vocal tract modeling from anatomical data, now carried out at MIT (Goldstein, 1979), should provide us with fresh insight in female, male, and child differences.

From Goldstein's still unpublished graphs of vocal tract outlines I have noted that the length of the pharynx measured from the glottis to the roof of the soft palate grows from 3.3 cm in the newborn child to 7.6 cm for the female aged 21 and 10 cm for the male aged 21. The length of the mouth measured from the back wall of the upper pharynx to the front teeth (alveolar ridge for the

newborn infant) grows from 5.5 cm for the newborn infant to 8 cm for the female of 21 and 8.5 cm for the male of 21. The tendency of relatively small variations of mouth cavity length with sex and age is more apparent than anticipated from earlier studies and would tend to minimize the range of "mouth cavity formant frequencies". The radical variations in relative pharynx length suggest that the relative role of front and back parts of the vocal tract could be reversed for a small child, i.e. that  $F_2$  of the vowel [i] would be a front cavity formant, whilst  $F_3$  is more dependent on the shorter back cavity. When front and back cavities are of more equal length, the dependency is divided and the  $F_3/F_2$  ratio smaller than for males, which is typical of females or children of an intermediate age.

#### The inverse transform

As noted already in the introduction, there has been a substantial amount of theoretical work directed towards the derivation of area functions from speech wave data. In practice, however, these techniques are limited to non-nasal, non-obstructed vocal productions and the accuracy has not been great enough to warrant their use in speech research as a substitute for cine-radiographic techniques. In the following section I shall attempt to comment on some of the main issues and problems. The usual technique, e.g. Wakita (1973), is to start out with a linear prediction (LPC) analysis of the speech wave to derive the reflection coefficients which describe the analog complex resonator. The success of this method is dependent on how well the losses in the vocal tract are taken into account. Till now the assumptions concerning losses have been either incomplete or unrealistic. Also the processing requires that the source function be eliminated in a preprocessing by a suitable deemphasis or by limiting the analysis to the glottal closed period. In spite of these difficulties the area functions derived by Wakita (1973; 1979) preserve gross features.

In general, a set of formant frequencies can be produced from an infinite number of different resonators of different length. We know of many compensatory transformations, such as a symmetrical perturbation of the single-tube resonator. However, if we measure the input impedance at the lips (Schroeder, 1967) or calculate formant bandwidths, we may avoid the ambiguities. A tech-

nique for handling tubes with side branches has been proposed by Ishizaki (1975).

According to Wakita (1979), the linear prediction method is capable of deriving an area function quantized into successive sections of equal and predetermined length providing the LPC analysis secures an analysis equivalent to  $M$  formants specified in terms of frequency and bandwidth.

An estimation of the total length and of the area scale factor require additional analysis data. An incorrect length estimate automatically generates compensatory changes in the area function which may be appreciable.

LPC analysis is a simple and powerful method of analysis but it fails in naturalness of representing the production process and as such is a poor substitute for a lossy transmission line representation. With the fresh eyes of a non-expert on the inverse transform, I would attempt to make the following suggestions. One is that  $M$  formants with associated bandwidths could have a greater predictive power than noted by Wakita. The area scale factor could be included in addition to the  $2M$  relative areas of his model. In general, with reservation for possible uniqueness problems,  $2M$  formant parameters - including bandwidths but not necessarily as many bandwidths as frequencies - would suffice for predicting  $2M$  independent area function parameters.

Thus, adding one more formant frequency to the  $M$  pairs of frequencies and bandwidths would suffice for estimating the total length of the  $2M$  system. Alternatively, from the  $2M$  formant measures, we could derive a model quantized into  $M$  equivalent tubes each specified by cross-sectional area and specific length, thus also predicting the total length, Fig. 15. The rationale for this reasoning is that all losses in the transmission line analogs are unique functions of the area and length dimensions. One could also design a three-parameter model of the vocal tract as in Fig. 15 with a constant larynx tube. The four parameters (lip parameter  $A_1/l_1$ ,  $x_c$  and  $A_c$ , and the total length) would hopefully be predictable from a specification of  $F_1$ ,  $F_2$ , and  $F_3$  and a bandwidth, say  $B_3$ , which appears to be more discriminating than  $B_1$  and  $B_2$ . If we omit the total length and sacrifice the bandwidth, we have approached the articulatory modeling of Ladefoged et al. (1978)

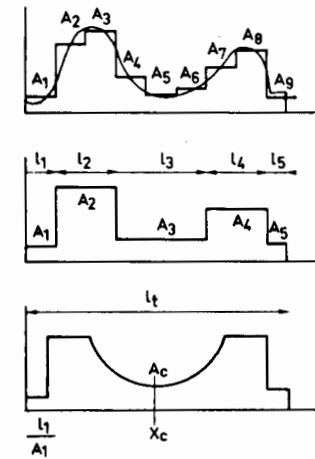


Figure 15. Continuous area function approximated by a constant larynx tube and 8 sections of equal length (top), by 4 sections of variable length and area (middle), and by a three-parameter model extended to include the total length (bottom). The constriction coordinate is zero at the lips.

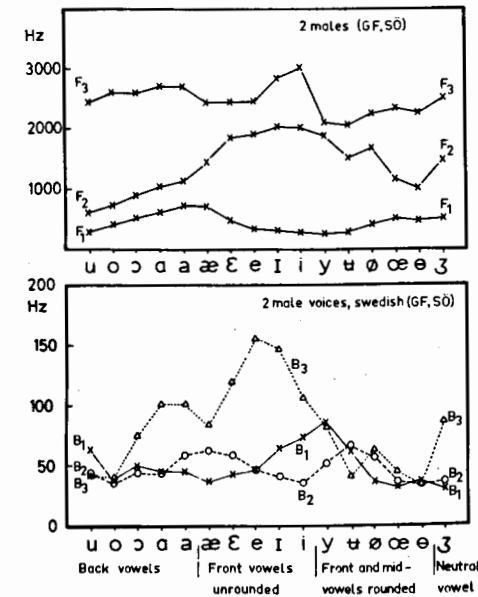


Figure 16. Frequency and bandwidth patterns of Swedish vowels (Fant, 1972).

which is based on correlational methods for deriving three articulatory parameters from  $F_1$ ,  $F_2$ , and  $F_3$ .

In general, bandwidths have less predictive power than frequencies. They are to some extent predictable from formant frequencies (Fant, 1972), Fig. 16. Furthermore, bandwidths vary with speaker, voice effort, and laryngeal articulations and are inherently difficult to measure.

Still, I do not want to rule out the use of bandwidths. The following examples may serve to illustrate their predictive power and limitations. First, a test of the uniqueness in predicting 2M area function parameters from 2M formant data. Take the simple case of  $M=1$  which implies a single tube resonator. What are the length and cross-sectional area of a tube with a specified first resonance frequency and bandwidth? The length is immediately given by  $F_1=c/4l$ . As shown in Fig. 17 the area is a single-valued function of bandwidth providing only one loss element is postulated (as in LPC analysis). If we include both the internal surface losses of a hard-walled tube and the radiation resistance, the bandwidth versus area attains a minimum at  $10 \text{ cm}^2$  and there are two alternative areas that fit the same bandwidth. The higher value could possibly be ruled out as being outside the possible range of human articulation. Similar ambiguities could also be expected in a more complex lossy transmission line model, as pointed out by Atal et al. (1978). However, one should note that their treatment of the invariance problem is not quite fair. They introduce more articulatory parameters than acoustic descriptors which obviously exaggerate the ambiguities. Next consider a two-tube approximation of the vocal tract, Fig. 18 (A), with a back tube of length 8 cm and area  $8 \text{ cm}^2$  and a front tube of length 6 cm and cross-sectional area  $1 \text{ cm}^2$ . The formant frequency pattern of  $F_1=275 \text{ Hz}$ ,  $F_2=2132 \text{ Hz}$ ,  $F_3=2998 \text{ Hz}$ ,  $F_4=4412 \text{ Hz}$  and all higher formants is exactly the same as that of a two-tube system with the same areas but the lengths reversed, i.e. a front tube of length 8 cm and a back tube of length 6 cm (Fig. 18 B). This length ambiguity rule is apparent from the expression for resonance conditions

$$\frac{A_2}{A_1} \text{tg} \frac{\omega l_1}{c} \times \text{tg} \frac{\omega l_2}{c} = 1 \quad (5)$$

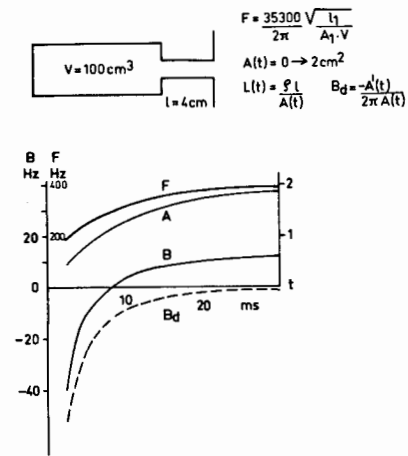


Figure 17. Bandwidth versus area of a single tube resonator taking into account internal losses and radiation load losses.

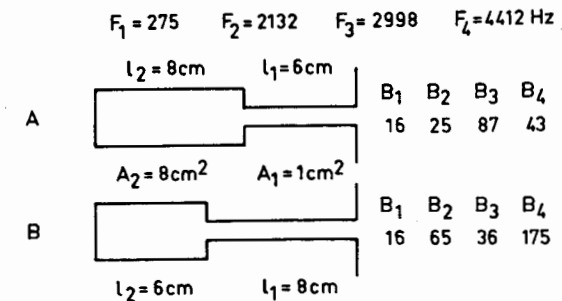


Figure 18. Two twin-tube resonators that provide the same F-pattern appropriate for the vowel [i], differing in terms of bandwidths.

If bandwidths are calculated taking into account both the interior surface losses and the radiation resistance by formulas given by Fant (1960), we find that  $B_2$  and  $B_4$  of Fig. 18 (A) are relatively low compared to  $B_3$ . In Fig. 18 (B),  $B_2$  and  $B_4$  are large compared to  $B_3$ . The different bandwidth patterns resolve the ambiguity. The physical explanation is that F2 and F4 of the first model are essentially determined by the back cavity and by the front cavity in the second model. The high damping associated with the surface losses in the narrow tube and the radiation resistance affect  $B_3$  of (A) and  $B_2$  and  $B_4$  of (B).

The two models do not differ in terms of  $B_1$ . Theoretically it would be possible to choose the correct  $l_1$ ,  $l_2$ ,  $A_1$ ,  $A_2$  of the two-tube model from a specification of  $F_1$ ,  $F_2$ ,  $F_3$  and either  $B_2$  or  $B_3$  or the ratio  $B_2/B_3$  or  $B_4$  or some combination of  $B_4$  and other bandwidths, e.g.  $(B_2+B_4)/B_3$ . In a real speech case the situation might be different if the glottal losses are large and execute high damping of the back tube resonances.

In practice it may take a ventriloquist to produce something similar to these two models. Possibly the one with a shorter back tube would fit into the vocal tract anatomy of a very small child, as suggested in the previous section.

In conclusion - to improve techniques for inferring vocal tract characteristics from speech wave data we need a better insight in vocal tract anatomy, area function constraints, and a continued experience of confronting models with reality - a balanced mixture of academic sophistications and pragmatic modeling.

#### References

- Atal, B.S., J.J. Chang, M.V. Mathews, and J.W. Tukey (1978): "Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique", JASA 63, 1535-1555.
- Fant, G. (1960): Acoustic theory of speech production, The Hague: Mouton (2nd edition 1970).
- Fant, G. (1965): "Formants and cavities", Proc.Phon.5, 120-141, Basel: Karger.
- Fant, G. (1966): "A note on vocal tract size factors and non-uniform F-pattern scalings", STL-QPSR 4, 22-30.
- Fant, G. (1972): "Vocal tract wall effects, losses, and resonance bandwidths", STL-QPSR 2-3, 28-52.
- Fant, G. (1973): Speech sounds and features, Cambridge, Mass.: MIT Press.
- Fant, G. (1975a): "Non-uniform vowel normalization", STL-QPSR 2-3, 1-19.
- Fant, G. (1975b): "Vocal-tract area and length perturbations", STL-QPSR 4, 1-14.
- Fant, G. (1976): "Vocal tract energy functions and non-uniform scaling", J.Acoust.Soc.Japan 11, 1-18.
- Fant, G. (1979): "Glottal source and excitation analysis", STL-QPSR 1, 85-107.
- Fant, G. and S. Pauli (1975): "Spatial characteristics of vocal tract resonance modes", in Proc. Speech Comm. Sem. 74: Speech Communication, Vol. 2, G. Fant (ed.), 121-132, Stockholm: Almqvist and Wiksell.
- Fant, G., K. Ishizaka, J. Lindqvist, and J. Sundberg (1972): "Subglottal formants", STL-QPSR 1, 1-12.
- Fant, G., L. Nord, and P. Branderud (1976): "A note on the vocal tract wall impedance", STL-QPSR 4, 13-20.
- Fant, G. and J. Liljencrants (1979): "Perception of vowels with truncated intraperiod decay envelopes", STL-QPSR 1, 79-84.
- Flanagan, J.L. (1965): Speech analysis synthesis and perception, Berlin: Springer (2nd expanded ed. 1972).
- Flanagan, J.L., K. Ishizaka, and K. Shipley (1975): "Synthesis of speech from a dynamic model of the vocal cords and vocal tract", Bell System Techn. J. 54, 485-506.
- Fujimura, O. and J. Lindqvist (1971): "Sweep-tone measurements of vocal-tract characteristics", JASA 49, 541-558.
- Goldstein, U. (1979): "Modeling children's vocal tracts", JASA 65, S25(A).
- Guérin, B., M. Mrayati, and R. Carré (1975): "A voice source taking into account of coupling with the supraglottal cavities", Rep. from Lab. de la Communication Parlée, ENSERG, Grenoble.
- Heinz, J.M. (1967): "Perturbation functions for the determination of vocal-tract area functions from vocal-tract eigenvalues", STL-QPSR 1, 1-14.
- Ishizaka, K., J.C. French, and J.L. Flanagan (1975): "Direct determination of vocal tract wall impedance", IEEE Trans. on Acoustics, Speech and Signal Processing, ASSP-23, 370-373.
- Ishizaki, S. (1975): "Analysis of speech based on stochastic process model", Bull. Electrotechn. Lab. 39, 881-902.
- Jospa, P. (1975): "Effets de la dynamique du conduit vocal sur les modes de résonances", Rep. de l'institut de phonétique, Université Libre de Bruxelles, 51-74.
- Ladefoged, P., R. Harshman, L. Goldstein, and L. Rice (1978): "Generating vocal tract shapes from formant frequencies", JASA 64, 1027-1035.
- Lindblom, B. and J. Sundberg (1969): "A quantitative model of vowel production and the distinctive features of Swedish vowels", STL-QPSR 1, 14-32.

- Lindqvist, J. and J. Sundberg (1972): "Acoustic properties of the nasal tract", STL-QPSR 1, 13-17.
- Mrayati, M. and B. Guérin (1976): "Etude des caractéristiques acoustiques des voyelles orales françaises par simulation du conduit vocal avec pertes", Revue d'Acoustique 36, 18-32.
- Mrayati, M., B. Guérin, and L.J. Boë (1976): "Etude de l'impédance du conduit vocal - Couplage source-conduit vocal", Acustica 35, 330-340.
- Nordström, P.-E. (1975): "Attempts to simulate female and infant vocal tracts from male area functions", STL-QPSR 2-3, 20-33.
- Öhman, S.E.G. and S. Zetterlund (1975): "On symmetry in the vocal tract", in Proc. Speech Comm. Sem. 74: Speech Communication, Vol. 2, G. Fant (ed.), 133-138, Stockholm: Almqvist and Wiksell.
- Schroeder, M.R. (1967): "Determination of the geometry of the human vocal tract by acoustic measurements", JASA 41, 1002-1010.
- Sidell, R.S. and J.J. Fredberg (1978): "Noninvasive inference of airway network geometry from broadband long reflection data", J. of Biomedical Eng. 100, 131-138.
- Sondhi, M.M. and B. Gopinath (1971): "Determination of vocal tract shape from impulse response at the lips", JASA 49, 1867-1873.
- Stevens, K.N. (1971): "Airflow and turbulence noise for fricative and stop consonants, static considerations", JASA 50, 1180-1192.
- Stevens, K.N. and A.S. House (1955): "Development of a quantitative description of vowel articulation", JASA 27, 484-493.
- Wakita, H. (1973): "Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms", IEEE Trans. Audio and Electroacoustics, AU-21, 417-427.
- Wakita, H. (1979): "Estimation of vocal tract shapes from acoustical analysis of the speech wave: the state of the art", IEEE Trans. Acoustics, Speech and Signal Processing, ASSP-27, 281-285.
- Wakita, H. and G. Fant (1978): "Toward a better vocal tract model", STL-QPSR 1, 9-29.

## DISCUSSION

Hisashi Wakita, Raymond Descout and Peter Ladefoged opened the discussion.

Hisashi Wakita: In determining the interrelationship between speech articulation and acoustics, we are particularly interested in the inverse problem, i.e. the estimate of vocal tract shapes from the acoustic waveform. There are various uncertain factors in deriving vocal tract area functions from the waveform, but it is an attractive method, because it is both the safest and easiest. (The problem with recent articulatory models for vocal tract shaping is that we do not yet know the exact parameters that control vocal tract shapes, in terms of articulators, and we do not have sufficient methods to obtain the data.) One of the most promising methods is the linear prediction (LPC) method, to estimate area functions from acoustic data. We do not know to what extent we can describe the details of the vocal tract shape, but by combining the LPC method with physiological data, we hope to improve this method.

One problem is the non-uniqueness, i.e. we can generate an infinite number of shapes having exactly the same frequency spectrum within a limited frequency band. To solve the uniqueness problem we have to impose constraints, physiologically determined constraints, or constraints determined by the higher harmonic structure. So far, the LPC method has been using formant frequencies and bandwidths, and in fact the final area function is sometimes quite sensitive to bandwidth. But we would like to get rid of bandwidth in the calculations: From the first three formant frequencies we can obtain the midsagittal view of the vocal tract, like in the Peter Ladefoged model, and to get at the unique shape of this midsagittal area function we may employ physiological constraints.

Another problem with LPC analysis is the vocal tract excitation and the losses, both within the vocal tract and at its boundaries, and these problems have to be solved in order to get more accurate vocal tract shapes. In fact, with the LPC method we can detect the closed glottis portion, where the interaction between sub- and supraglottal cavities is minimized, which makes for more accurate area functions. A further draw-back of LPC is

that we have to start from the very simple assumptions of a simple loss at the glottis and a lossless acoustic tube. On the other hand, you can make a production model as complex as you wish, - you can add any realistic losses along the vocal tract or at the glottis that you like, but as analysis model there is a strong limitation in incorporating losses and other factors. So at this moment, the imminent problem is how to attack the loss problems and the source uncertainties.

Raymond Descout: Very little original data has accumulated on area functions, because collecting it is difficult, from a technical point of view. On the other hand, deriving vocal tract area functions from acoustic data has some disadvantages: with LPC techniques we only get pseudo area functions, and with acoustic measurements, which I previously worked on, there is a great problem in dynamic measurements, especially. Further, interest has largely centered on the midsagittal view of the tract, but we need information about the frontal view as well, which may be obtained with the new techniques of computerized tomography. We need this information in order to turn the midsagittal view into a three-dimensional area function, and to determine the shape factors that are necessary for the introduction of losses in our models.

All the articulatory models proposed are based upon vowel configurations, and when we try to make dynamic simulations on the articulatory model, everything that we do not know about the consonants is put into a special coarticulation and transition rule. We need more information on the consonants.

The acoustic model of the vocal tract is derived from the propagation equations, based on assumptions of symmetrical, equal length sections, - but to do an inverse transform you really need a very appropriate model which includes the shape factors that are necessary for the loss calculations, because the mathematical technique involved in the transformation is stupid in the sense that the result will be adjusted according to mathematical criteria, but this may not result in a realistic vocal tract. Therefore, I think that doing inverse vocal tract transforms is premature: we must work first of all on the proposition of the best production model, including shape factors and losses, before trying to do inverse vocal tract transforms.

Due to the progress made in articulatory modelling and to the limitations of LPC-techniques, we have witnessed a come-back of studies on vocal tract and vocal source simulations. To refine the articulatory model, we need further physiological data.

In conclusion: I do not think that LPC will give us a better understanding of speech production (it is, however, excellent for synthesis purposes). We need more studies on the relationship between articulatory parameters / area functions / vocal tract shapes.

Peter Ladefoged: Gunnar Fant showed us many years ago that what is important in characterizing speech are the first three formant frequencies, and you can even get a great deal of a speaker's personal quality with just three formant frequencies. But with the inverse transform, to get as far as eight tubes (which is only a coarse model of the vocal tract), you need at least four formant frequencies and their bandwidths, and with eighteen tubes you need nine formant frequencies and bandwidths, etc. Now something is wrong here: any phonetician can draw, more or less accurately, the midsagittal view of a given speaker's vowels, and we ought to be able to develop an algorithm that will go from the acoustics to the tract shape. There are of course problems - we do not actually observe the tract shape, only the midsagittal dimensions, and there are only very limited sets of data that tell us how to derive the tract shape from the sagittal dimension.

The work of Lindblom and others has shown that you can produce an [i:] with your jaw in a more or less open position, i.e. one has the ability to control tract shapes using different articulatory procedures, and it is of great interest to us to know how we exert that control and less interesting what the muscles do. Eventually, we have got to be able to go from acoustic structures, finding out what the tract shape is, and then deducing from that what the underlying control signals must have been.

Gunnar Fant: I agree with the main points of the discussants. Inverse transforms cannot make up for our great lack of physiological reference data.

My suggestions for improving inverse transform techniques

in part supported by the previous discussions are: (1) we should model the vocal tract in terms of lossy transmission line sections instead of the simplified LPC model, (2) we should not expect to generate a larger number of independent production parameters than we have independent and well specified speech wave descriptors relating to the vocal tract transfer function. Overspecified area functions are necessarily non-unique, whereas a balanced specification can be, but need not be, unique. With proper model and parameter constraints, a 32-section area function model may be generated from a set of 3-6 articulatory parameters and controlled by the same number of acoustic parameters. It remains to be seen if we can extract more than four independent acoustic parameters. (3) The vocal tract total length should be derivable from one extra independent acoustic parameter.

Our discussion concerning bandwidths is still rather academic and we appear to share a doubt concerning the specificational value of bandwidths. Theoretically the set  $F_1 F_2 B_1 B_2$  could suffice to specify a three-parameter model extended with a fourth parameter, e.g. the total length. This might hold for a resonator model only but not for a true vocal tract with less predictable bandwidth sources and the limited accuracy in bandwidth measurements. A more efficient set of acoustic parameters would be  $F_1 F_2 F_3$  and  $B_3$ . From my Fig. 16 illustrating bandwidths of Swedish vowels it is seen that  $B_3$  is a good correlate of degree of lip opening and also mouth opening. However, vowel bandwidths including  $B_3$  are to a high degree predictable from formant frequencies. The role of bandwidths in an LPC model is not the same as that of a true vocal tract model. This is an important distinction. The LPC bandwidths, e.g.  $B_3$ , may come out quite different from those of real speech or from simulations by an improved model. The bandwidths we need for the inverse LPC based transforms are the bandwidths of a production model which has losses at the glottis only and locks the cavity wall shunt. From the true formant frequencies and bandwidths we thus have to make a best guess of what bandwidths the LPC model would generate. This is in the line of the recent work of Hisashi Wakita (1979).

Kenneth Stevens: With regard to what a male speaker does in order to compensate relative to the [u:] of a female: if we define narrow vowels as having so narrow a constriction that turbulence is just not generated, is it conceivable then that males, who generate a greater air flow than women, cannot round the vowels as much as can women, and therefore the formants are not lower than those of women?

Gunnar Fant: It could be, but in Swedish the vowel [u:] as well as [i:], [y:], and [ɤ:] are generally produced, by males and females alike, with a diphthongal glide passing through a relatively constricted phase in which some turbulence may be generated. I would rather expect different male and female articulations to be aimed at some criterion of perceptual invariance of which we do not know too much yet.

Antti Sovijärvi asked Gunnar Fant what his concept is about nasalized vowels.

Gunnar Fant: An essential characteristic of nasalization independent of the specific resonances added is the reduced F1 amplitude which is especially apparent in an oscillographic analysis. What appears to be a sub-F1 nasal formant is often a voice source feature which is relatively re-inforced because of the F1 reduction.

Hisashi Wakita: As long as the calculations are based on the first few formant frequencies, the problems in inverse transformation are rather equivalent with different methods. To uniquely determine a six tube vocal tract shape, LPC uses the first three bandwidths. If you want a smooth area function, you have to specify one of the higher frequency characteristics, and to do that you have to impose some kind of constraint, which is what Dr. Ladefoged does. And whatever the method, if you do not want to use bandwidth, you have to use some other kind of information to uniquely determine the spectra, and any information will do as long as you are able to reconstruct the original spectrum with its original bandwidths - so bandwidth is in fact a very important parameter.



Gunnar Fant: It would be interesting to see how far you would get if you started out with F1, F2, and F3 and then predicted B1, B2, and B3 from the formulas that I have.

Peter Ladefoged: I have tried using Hisashi Wakita's formulae with Gunnar Fant's type of predicted bandwidths (and other bandwidths from the literature), and it did not work, - I got absolutely impossible vocal tract shapes. Regarding Atal's vocal tract shapes that produce identical formant frequencies: some of them are quite impossible, the tongue just cannot produce some of those shapes.

John Holmes: I wish to emphasize the difficulty of mathematically deriving the vocal tract from the speech waveform, because we know too little about the glottal source. Gunnar Fant emphasized that the closed glottis portion is better suited than the open glottis portion to work out the supraglottal characteristics, but (as can be seen on the Farnsworth vocal chord movie of about 1940 and from Tom Baer's work), even when the vocal chords are closed there is sufficient ripple and surface movement for there to be an effective volume velocity input into the vocal tract, which means that your resultant waveform is never a force-free response, - and this is one of the things that makes bandwidths so difficult to estimate, because it is quite possible that ripple in vocal chord surface could actually be causing the formant amplitude to be still building up even, in exceptional cases, during the closed glottis period. I think this supports the view that we have to work from much more basic information and use articulatory constraints rather than to derive vocal tracts by purely mathematical techniques from some artificial and unrealistic production model.

Gunnar Fant: I can only agree with your statements. It is necessary to learn more about the human voice source in order to improve our methods of inverse transforms.

Osamu Fujimura: We can obtain cross-sectional vocal tract shapes with the regular computerized tomography, but only at great costs, because the X-ray dosage is tremendously high, a requirement of brain diagnoses that demand a very good density solution.

But I think the machine can be adjusted and the X-ray dosage reduced for our purposes, where we are really only interested in the distinction between matter and air.

Mohan Sondhi at the Bell Laboratories has proposed an acoustic impedance measurement using an impulse-like excitation at the lips, which can give us complete information about the area function of the vocal tract, because we obtain two sets of infinite series, i.e. the poles and the zeroes of the impedance function that together uniquely determine the vocal tract shape, without having to assume or measure losses. I think that there is one major difficulty with this technique: the subject articulates silently, i.e. he has no auditory feed-back, and we cannot be sure about the actual gestures. That problem can be overcome if we simultaneously monitor the vocal tract with e.g. the X-ray micro-beam method.

Gunnar Fant: The micro-beam system will certainly provide us with excellent data about speech articulation, but will it provide us with all the details that we want about the vocal tract, like the exact dimensions of the pharynx and larynx cavities?

Osamu Fujimura: We can obtain data on cross-sectional shapes, because we can place pellets also outside the midsagittal plane, - the only constraint being that we cannot use too many pellets at the same time, which will increase the X-ray dosage, but it is not easy to place pellets on the pharyngeal walls, which is a limitation of the method. However, we have a new stereo-fiber-scope which can be used for three-dimensional optical observations of the pharynx, and I hope in the future to be able to develop a technique that will supplement the X-ray technique with this kind of optical information.

Raymond Descout: I am presently working with a prototype CT (computerized tomography) scanner, which scans in five seconds, and we are trying to lower the X-ray dosage to ten percent the normal dosage, because all we need is to see the difference between air and flesh. There is still a problem with the CT technique, though, and that is determining exactly the position of the slice relative to the skin and the rest of the person.

## MODERN METHODS OF INVESTIGATION IN SPEECH PRODUCTION

Osamu Fujimura, Bell Laboratories, Murray Hill, New Jersey 07974

Chairperson: Celia Scully

1. Descriptive Theory and Modeling of Speech Production

The process of speech production involves many aspects which may be treated by different disciplines of science. As much as we deal with speech as signals representing linguistic codes, it is clear that we need to have a descriptive framework of the linguistic message, so that we can relate the observed physical phenomena to the units that are used in the codes. Both segmental and supra-segmental specifications have to be given, as well as appropriate indications of surface syntactic (and semantic) information.

In addition to the lexically distinct accentual patterns and different intonational patterns for phrase structures, modulations of voice pitch and duration may be extensively used in conversational speech reflecting, e.g., focus, emphatic contrast, contextual and statistical predictabilities of the word, etc. Since speech phenomena always involve paralinguistic factors, such as the speaker's emotional state and idiosyncrasy, a way of describing those is also needed; or at least we must have a clear idea about what relevant factors have to be kept constant to make the comparison of different linguistic units meaningful. These considerations become more and more important, as we make progress in speech research. There are some emerging efforts in this direction, both in theory and experiment. The metric theory (Liberman and Prince 1977) for description of stress and intonation patterns of English constitutes a good example of such theoretical progress in this area, and a pitch contour synthesis-by-rule experiment based on this theory (Pierrehumbert 1979) suggests rapid progress in this field.

The notion of segments is also being revisited in connection with the significance of larger segmental units. The basic idea is to concatenate segmental units, whether phonemes, syllables, phonological words or phrases, to form larger units, and give suprasegmental modulations as patterns assigned to the larger units. Experiments in synthesis by rule attempt to evaluate models of this process. The notion of temporal modulation can be clarified only by referring to a well-defined model of speech dynamics that im-

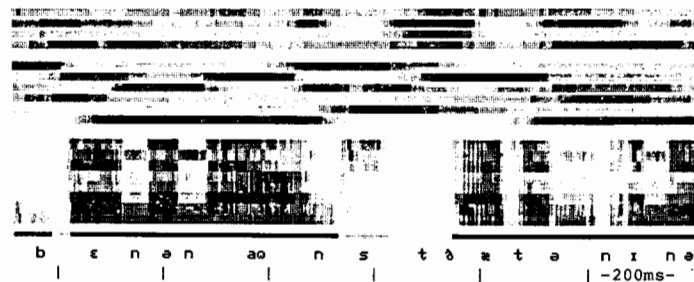
plements an abstract specification of concatenated strings of units. Such a phonetic realization process would be characterized by different dynamic (i.e. temporal) characteristics for individual articulators, and the realized phonetic events corresponding to the so-called (phoneme size) "segments" are in general not in synchrony. Therefore, discontinuities observed in acoustic signals, such as the voice onset, stop release, etc., may not reveal some of the important aspects of the temporal characteristics of speech.

Gunnar Fant (1962) described a fine subsegmentation of acoustic signals based on their apparent discontinuities and interpreted such spectrographic representations of speech in terms of overlapping acoustic properties, roughly similar to, but crucially different from, the linguistic distinctive features.

In order to account for the full information contained in speech signals and its human perception, one has to go well beyond this basic sketch. The spectral modulation of the speech signal is in one aspect discontinuous and in the other continuous. This dual nature of speech may be seen most obviously when we compare a gross spectrographic representation with an articulatory representation (see Figure 1) (Miller and Fujimura 1979). This qualitative difference between articulatory movement and its consequent acoustic temporal pattern stems from the inherent non-linearity between the two levels of speech representation.

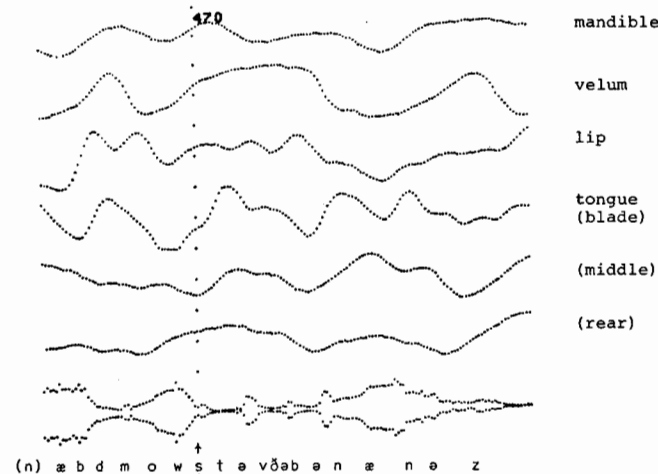
Recent studies are revealing interesting details of articulatory processes in relation to the phonological structures of the message. It is being shown that a simple model of concate-

FIGURE 1



A combined articulatory-acoustic representation of part of a sentence 'Ben announced that an innocent-seeming infant had nimbly nabbed most of the bananas', uttered by a male native speaker of American English (Fresno, California). The upper part pertains to pellet positions, as obtained by the computer-controlled x-ray microbeam system, and the lower part a simplified (8 frequency-band) spectrographic pattern. In the lowest horizontal line black, gray, and white represent, respectively, voiced, voiceless, and silent states of the speech signal, and the phonetic symbols underneath are selected and placed automatically based on the articulatory information as well as the voicing state of the sound. The articulatory gesture is represented by the topmost 4 stripes for front (dark)/back (light) movements of the pellets placed on (from top) the lower lip, the blade, mid and rear portions of the tongue, and below these by the 6 stripes for up (dark) - down (light) movement of (from top) the lower lip, the mandible, the three parts of the tongue, and the velum (dark for low) (see Nelson [1979], Miller and Fujimura [1979]).

FIGURE 2



Time functions representing vertical movements of the 6 pellets (the same material as in Fig. 1). The lowest trace depicts the speech waveform envelope. The arrow in line with a vertical array of dots is placed at the beginning of the voiceless segment for /st/.

nating phoneme-size units into larger phonological units, taking care of "coarticulation" phenomena by smoothing the movements, simply does not work. This is so particularly because within each syllable (or more exactly syllable core, see Fujimura and Lovins (1978), Fujimura (1979b)) there is something much more ad hoc about the temporal structure of phonetic events as syllabic ingredients. Such ad hoc characteristics are largely dependent on the language (and dialect) and therefore cannot be specified by a universal phonetic principle. By examining articulatory processes for relevant organs in movement, allowing for different dynamic characteristics and freedom of asynchrony in motor control for different articulatory (or phonatory) dimensions, we can obtain some insight into the nature of the temporal organization of phonetic events (Fujimura, forthcoming-a). Even inversions of temporal relations of peak activities for individual articulatory gestures are observed, from a phoneme string point of view. For example, as shown in Fig. 2, the syllable /mowst/ in a sentence utterance shows that the labial constriction for the glide /w/ manifests its peak activity during the voiceless period for /st/ toward the articulatory closure of /t/. A general principle governing phonetic structures of syllables (for the language) guarantees this looseness of temporal ordering within the syllable core to be irrelevant for phonological identification of this form (see Fujimura and Lovins (ibid)).

A useful descriptive framework thus seems to be one based on individual articulatory events related to elementary (functional) features of the syllable core. Basic notions, such as concatenation, coarticulation, assimilation and dissimilation have to be revisited quantitatively in light of such a descriptive model. It is time for us to produce experimental evidence for or against specific intuitive predictions. The scope of such experimental work is now being drastically expanded, thanks to newly available tools. It must be emphasized, however, that any of the available techniques for physical measurement, even in the future, is not likely to provide us with a complete picture of the physiologic/physical phenomena of speech production by itself. In order to interpret the results of measurements at different levels and relate them to each other, which is the task given to speech scientists for understanding the speech production process, we need to devise some new tools. Computational models of the natural speech

production apparatus are being studied as such tools. For example, a three-dimensional static model of the tongue has been constructed using the finite element method (Kiritani et al. 1976) and is being used for studies of control characteristics of vowels (Fujimura and Kakita 1979).

A quantitative study of the gesture for the vowel [i], based on the tongue model, has suggested that the contraction of the posterior portion of the genioglossus muscle alone can give rise to a reasonable shape of the tongue and a consequent formant pattern for this vowel, but a slight deviation from the correct magnitude of contraction would cause quick deviation from the acceptable phonetic value. On the other hand, if we use a set of muscle components, in conformity with available electromyographic findings, we find such sensitivity to the degree of contraction is eliminated and the resultant phonetic quality becomes very stable and easy to achieve with a wide latitude of physiologic control. This points to the question of the quantal nature of speech as proposed by K. N. Stevens (1972), and also to the significance of feedback in different situations of speech production, including normal and artificial (such as the bite block) circumstances. With respect to the quantal nature, it seems that the crucial issue is the choice of the input level, at which the change of the controlled quantity in question is compared with that at the output, i.e. the acoustic characteristics such as formant patterns. The midsagittal tongue contour or a parametrically represented area function does not seem to be the correct input to the system for this specific discussion. The three-dimensional structure of the tongue combined with its volume incompressibility seems to play an essential role in characterizing the nonlinearity of the input-output mapping. Also, if, as our tongue model study seems to suggest, what is important in achieving a phonetic goal of articulatory gesture is selecting the pertinent set of muscles (with a certain balance of relative activities) rather than the exact magnitude of muscular contraction (excessive contraction resulting only in more or less unaffected physical consequences) the observed robustness of articulation under affected conditions seems more readily explicable than we had thought before. Gross orosensory feedback information also seems to play an important role in this connection (Perkell 1977).

Within the hierarchy of the natural process of speech production, the higher the level, the less applicable direct physical measurements are. Recent efforts by psychologists (see e.g. Sternberg et al., (1978)) are focused on temporal aspects of motor control, in the attempt to infer basic mechanisms of cortical programming and its execution. Studies of highly skilled performances in nonspeech areas seem to point to the understanding that in routine human actions the temporal course of a physical state takes a fixed preprogrammed pattern. In speech, articulatory events are decomposable into elementary gestures, such as lip movements for bilabial stops and velum raising for nasal-to-non-nasal transitions. Recent articulatory measurements indicate relatively constant speeds of such movements in a wide range of conditions when influences of certain separable factors are excluded (see the co-report on speech production by Sawashima (vol. I, p. 49-56)).

It has been argued (MacNeilage 1970) that the notion of invariant gestures for phonetic units is untenable in consideration of the high number of different contextual conditions. Such estimates, however, customarily depend on phoneme-size phonetic units as the basis of assuming targets. Based on an analysis that syllables are separable into cores and phonetic affixes, and each core into relatively constant dynamic patterns of initial and final demisyllables (the latter including the central portion of the syllable), we can actually construct for English a complete inventory of phonetic (concatenative) segmental units that contains less than 1,000 items for virtually all possible English phonetic forms (Lovins et al., 1979). Assuming that each inventory item is given phonetic indexes (syllable features) representing articulatory gestures, and also temporal parameters that are sensitive to nonsegmental conditions such as stress/accent, speed of utterance, etc., it does not seem implausible that the human brain can store all necessary phonetic patterns in the given language. An experimental evaluation of this new view is being attempted by synthesis-by-rule experiments using a demisyllabic inventory. A concrete model of acoustic realization of syllable features is being studied by Mattingly (1977). The psychological reality of the core-affix decomposition as well as the syllable itself is still to be examined.

## 2. Physiological Studies - Muscle Controls

The study of the physiology of speech production has seen remarkable progress in the past decade, even though there are still many unsolved basic questions. One general question is which muscle plays the principal role of implementing motor commands for a given phonetic gesture, viz. an elementary articulatory event. Electromyographic studies have revealed, for example, that the glottal abduction reflecting the devoicing gesture is related to the activity of the posterior cricoarytenoid muscles, whereas glottal adduction is achieved by several different muscles, including the interarytenoids, in varied ways depending on linguistic (and paralinguistic) functions (Hirose and Gay 1972; Hirose et al. 1978).

Hirano recently studied the anatomy and physiology of the vocal cords using various advanced techniques such as electron microscopy, histochemistry, electromyography, electric nerve stimulation, high speed motion picture, mechanical measurements, applied to both human and animal larynges (Hirano 1977). He arrived at an approximation of the complex anatomical structure by two (or three) loosely coupled parts, viz. cover and body. The cover seems to be responsible for the major part of the vibratory movement, showing large three-dimensional excursions, whereas the body contains the so-called vocalis muscle and participates in active parametric control of the vibrating system (see also Fujimura (1979a)). Baer (1975) has contributed a detailed study of excised canine larynges, and Titze and Talkin (1979) are contributing a new computerized model of the vocal cord vibration process.

Pitch control is an important topic from both lexical distinction and sentence-intonation points of view. The physiologic mechanism is not completely understood, but much is known now about the function of the cricothyroid muscle in relation to the voice fundamental frequency. There are cases where the voice fundamental frequency does not reflect the phonological accentual pattern because of the interaction between the consonantal control of voicing/tenseness and the vocal fold vibration frequency, but the electromyographic signal of the cricothyroid does (see Fujimura (forthcoming-b)).

Lingual muscles are difficult to study even with the best available electromyographic techniques because of the complex

interdigitation of a number of muscles forming the main body of the tongue. Nevertheless, the rather limited information obtained by EMG measurements are indispensable in inferring muscular functions relative to specific phonetic gestures.

Controlled interference by such techniques as anesthesia and bite block, has been experimentally induced, in order to evaluate the roles of feedback loops in speech production (Lindblom et al. 1977). In real utterance situations, mandible height is not necessarily correlated with tongue height either positively or negatively. For example, for the American English vowels /e/ and /ɛ/ in sentence utterances, we have found in our X-ray microbeam data that a tongue height measure does distinguish occurrences of the two vowels very clearly, but that mandible height can be either lower or higher for one vowel than the other. Mandible height seems to reflect the stress status of the vowel, serving a function that is partially independent of the vowel height specification.

### 3. Physical States of Organs

Neural control of the larynx is parametric in the sense that gross average states of the larynx rather than details of vibratory changes of the peripheral shapes of the vocal cords are adjusted. For this reason, if we measure the laryngeal state during an utterance, the measurement may be taken at a relatively slow sampling rate such as 50 samples/second and averaged over a period like 20 msec. The fiberoptic technique developed at the University of Tokyo is appropriate for this purpose (Sawashima and Hirose 1968).

There have been successful studies of segmental control, such as manners of consonantal articulations in different languages (see for a review, Fujimura (1979a)). Here again, there are cases where the acoustic signal cannot answer a question about control. The laryngeal maneuver for pitch control seems related to vertical movements of the larynx as well as other gross appearances of the glottal area, and this may give us an opportunity to learn about pitch control even for devoiced syllables. A recent improvement of the fiberoptic has made it possible to record two images side by side on the film stereoscopically, so we can measure the distance between the objective lens and the object (Fujimura et al. 1979). For many phonetic studies on qualitative states of the

glottis, on the other hand, electric resistance measurements are being used as a readily applicable tool (Fourcin 1977, Frøkjær-Jensen 1968). Characteristics of voice source signals have gained renewed interest. Gunnar Fant (this volume, p. 79-108) is contributing a new insight about the interaction between the source and the vocal tract by closely examining speech waveforms. Flanagan et al. (1975) used their two-mass model of the vocal cords for simulating turbulence generation in the coupled source-vocal tract system.

The lips are obviously the easiest object to measure among different articulators, particularly with the use of a powerful stroboscopic technique (Fujimura 1961). A modern computerized system for measurement of the lips and mandible positions as well as linguapalatal contact is now available at the University of Alabama (McCutcheon et al. 1977). A servomechanistic technique can be used for a more general analysis of the natural articulatory systems such as the mandible and the lips. Such a measurement system has been implemented at the University of Wisconsin, Madison, and the control mechanisms of the lips are being studied assuming a linear system with feedback loops (Muller and Abbs 1979). The frequency response of such looped systems seems to allow actively controlled movements of visco-elastic systems via brainstem feedback for the majority of speech events. It should be emphasized, however, that the peripheral parts of articulators do not necessarily move together with the neurally controlled body of the same organ, and it is the former that determines acoustic consequences.

Dynamic characteristics of articulators in speech have been a vital issue in speech research. Several interesting proposals have been made about the basic principle of articulatory gestures trying to relate abstract and discrete phonological codes to the temporal structures of continuous speech phenomena (see Kent and Minifie (1977) for a review). Information on actual movements of the principal organs, in particular the tongue, is badly needed for such a study. Relatively large amounts of data obtained from the same subject are necessary to cope with an inherent variability of speech production phenomena. Collection of comparable data from many subjects, wherever possible, is another necessity for understanding the other aspect of human variability.

There are several methods that have been proposed and tested for observing tongue movements. Dynamic palatography (Fujimura et al. 1973b) represented an early attempt to computerize tongue observation for acquisition and processing of large amounts of data. It is also being applied to training of children in speech and hearing clinics in Japan. Other more recently proposed techniques include optical distance measurement between selected points on the palate and the nearest tongue surface. Magnetic (Sonoda 1977) as well as ultrasonic (Minifie et al. 1971) measurements also have been proposed.

The most direct and informative method of observing tongue movement is the use of X-rays for lateral views of the tongue. There used to be two factors that made radiographic measurements impractical for obtaining a large quantity of speech data. One is the radiological disturbance given to the subject. For this reason the exposure had to be limited usually to one or two minutes total per subject. The tedious and inefficient frame-by-frame analysis of the photographic images constituted another problem. The computer-controlled X-ray microbeam system was devised precisely to overcome these difficulties (Fujimura et al. 1973a). A full-scale system is now in operation at the University of Tokyo (Kiritani et al. 1975), and is producing useful results.

Several metal pellets are placed on selected points on the tongue and other articulators, usually but not necessarily in the midsagittal plane. A computer directs a thin X-ray beam to search around a predicted position, for each pellet, based on its past position and movement, verifies the current position, and repeats the procedure to look for the next pellet. By the combination of high sensitivity of the X-ray detector and an efficient use of the given total dosage for determining pellet positions, without exposing any unnecessary portions of the body for the specific purpose, the total radiographic exposure is incomparably smaller than that which would be used by film recording with an image-intensifier. The pellet position at each sample time, typically every 10 ms or less for 6-8 pellets, is digitally stored in the computer memory in real time. The experimenter, and the subject if desirable, can monitor the detected pellet movements. Powerful computer programs have been designed and implemented at Bell Laboratories in order to give the experimenter an efficient

tool for interactive data analysis. Figure 1 represents one of the results, including an automatic annotation of the speech material with phonetic symbols (Nelson 1979).

An independent estimation of area functions by acoustic input impedance measurement has been proposed (Sondhi and Gopinath 1972). There is a nontrivial mapping process between the acoustically effective area function and the state of the speech organs (Mermelstein 1973). On the other hand, the so-called pseudo-area function that is conveniently derived by the well-established linear-prediction coding scheme (LPC) is not a true representation of the vocal tract characteristics proper (see Fant, p. 79-108, and Wakita, p. 151-172 (this volume)). Therefore, it is very desirable to have such independent measurement of the true area function, particularly if a simultaneous X-ray observation can be made for direct comparison of tongue shape (pellet positions) and the effective area function. The use of the recently developed CAT technique is also being attempted for static gestures.

#### 4. Statistical Processing of Production Data

The availability of a large amount of production data encourages researchers to use advanced techniques of statistical processing of data such as multidimensional analysis (INDSCAL (Carroll and Chang, 1970) or PARAFAC (Harshman et al. 1974)), as well as principal component analyses. Through purely statistical processes, constituent (static) gesture components have been derived from both hand-traced midsagittal contours of the tongue of many speakers (Ladefoged 1977) and automatically tracked pellet position data for each of a few speakers (Kiritani and Imagawa 1976). These inductive methods give us purely phenomenologically derived "phonetic coordinates" for describing articulatory characteristics of a class of phonetic units, which is defined by the particular choice of the speech material used for this data processing. It is an intriguing question to ask if we can have a universal descriptive framework that explains the relations between different aspects of categorization of phonetic units (see Ladefoged's co-report on speech production (vol. I, p. 41-47)).

The use of multiple regression technique (both linear and nonlinear) must be mentioned in connection with the inverse mapping from acoustic characteristics to articulatory conditions. In addition to the more traditional method of analysis-by-synthesis,



which also is being used extensively (Fujisaki 1977), such new computational means seem to promise a new trend of research. Multiple regression techniques have been used for interpreting both durational parameters (Liberman 1978), and articulatory data (Nakajima 1977, Shirai and Honda 1977). The former used automatic processing of reiterant speech signals (Liberman and Streeter 1978), having the subjects mimic a sentence by a repetition of the same syllable, such as [ma], and attempted a best match between model-predicted and measured syllable durations by adjusting relative contributions of different phonologic and syntactic factors. The latter, using nonlinear regression, assumes a simple dynamic model of the physical movements of the articulators to determine the parameters that characterize such a physical system.

#### 5. Concluding Remarks

When we define a domain of problems, such as normal speech, speech of a particular speaker, vowels as opposed to consonants, phonology as opposed to syntax, etc., we always need some understanding of the problems surrounding that domain. By knowing what happens just outside the boundary of the domain of immediate interest, in accordance with the principle of continuity, we always gain better insight as to how to delimit the domain. Thus, for example, speech pathology is another intriguing area of phonetic research. Needless to say, we would like to learn how people perceive speech, in order to investigate how people speak, because the real-life speech behavior is always a continuous mixture of production and perception.

#### References

- Baer, T. (1975): Investigations of phonation using excised larynxes, PH.D. dissertation, M.I.T.
- Carroll, J.D. and J. Chang (1970): "Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition", Psychometrika 35, 283-319.
- Fant, G. (1962): "Descriptive analysis of the acoustic aspects of speech", Logos 5, 3-17.
- Flanagan, J.L., K. Ishizaka, and K.L. Shipley (1975): "Synthesis of speech from a dynamic model of the vocal cords and vocal tract", Bell Syst. Tech. J. 54, 485-506.
- Fourcin, A.J. and E. Abberton (1977): "Laryngograph studies of vocal-fold vibration", Phonetica 34, 313-315.
- Frøkjær-Jensen, B. (1968): "Comparison between a Fabre glottograph and a photo-electric glottograph", Annual Report of the Institute of Phonetics, University of Copenhagen 3, 9-16.
- Fujimura, O. (1961): "Bilabial stop and nasal consonants: A motion picture study and its acoustical implications", JSHR 4, 233-247.
- Fujimura, O., S. Kiritani, and H. Ishida (1973a): "Computer controlled radiography for observation of movements of articulatory and other human organs", Comput. Biol. Med. 3, 371-384.
- Fujimura, O., I.F. Tatsumi, and R. Kagaya (1973b): "Computational processing of palatographic patterns", JPh 1, 47-54.
- Fujimura, O. and J. Lovins (1978): "Syllables as concatenative phonetic units", in Syllables and segments, A. Bell and J.B. Hooper (eds.), 107-120.
- Fujimura, O. (1979a): "Physiological functions of the larynx in phonetic control", in Current issues in the phonetic sciences (Proc. of the IPS-77 Congress, Miami, Florida, Dec. 17-19, 1977) vol. I, 129-164, H. and P. Hollien (eds.), Amsterdam.
- Fujimura, O. (1979b): "An analysis of English syllables as cores and affixes", Zs.f.Ph., Sign and system of language, Heft 4/5, 452-457.
- Fujimura, O., T. Baer, and S. Niimi (1979): "A stereo-fiberscope with a magnetic interlens bridge for laryngeal observation", JASA 65, 478-480.
- Fujimura, O. and Y. Kakita (1979): "Remarks on quantitative description of the lingual articulation", in Frontiers of speech communication research, S. Ohman and B. Lindblom (eds.), 17-24, London: Academic Press.
- Fujimura, O. (forthcoming-a): "Elementary gestures and temporal organization -- What does an articulatory constraint mean?", Proc. of the International Symposium on the Cognitive Representation of Speech in their series 'Advances in Psychology', G. Stelmach and P. Vroom (eds.).
- Fujimura, O. (forthcoming-b): "Fiberoptic observation and measurement of vocal fold movement", Paper presented at the Conference on the Assessment of Vocal Pathology, NIH, Bethesda, Maryland, April 17-19.
- Fujisaki, H. (1977): "Functional models of articulatory and phonatory dynamics", in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.), 347-366, Tokyo: University of Tokyo Press.
- Harshman, R., P. Ladefoged, L. Goldstein, and J. Declark (1974): "Factors underlying the articulatory and acoustic structure of vowels", JASA 55, 385.
- Hirano, M. (1977): "Structure and vibratory behavior of the vocal folds", in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.), 13-30, Tokyo: University of Tokyo Press.



- Hirose, H. and T. Gay (1972): "The activity of the intrinsic laryngeal muscles in voicing control -- an electromyographic study", Phonetica 25, 140-164.
- Hirose, H., H. Yoshioka, and S. Niimi (1978): "A cross language study of laryngeal adjustment in consonant production", University of Tokyo AB RILP 12, 61-72.
- Kent, R.D. and D. Minifie (1977): "Coarticulation in recent speech production models", JPh 5, 115-133.
- Kiritani, S., K. Itoh, and O. Fujimura (1975): "Tongue pellet tracking by a computer-controlled X-ray microbeam system", JASA 57, 1516-1520.
- Kiritani, S. and H. Imagawa (1976): "Principal component analysis of tongue pellet movement", University of Tokyo AB RILP 10, 15-18.
- Kiritani, S., F. Miyawaki, O. Fujimura, and J.E. Miller (1976): "A computational model of the tongue", University of Tokyo AB RILP 10, 243-251.
- Kiritani, S., S. Sekimoto, and H. Imagawa (1977): "Parameter description of the tongue movements for vowels", University of Tokyo AB RILP 11, 31-38.
- Ladefoged, P.N. (1977): "The description of tongue shapes", in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.), 209-222, Tokyo: University of Tokyo Press.
- Liberman, M.Y. (1978): "Modeling of duration patterns in reiterant speech", in Linguistic variation, models and methods, D. Sankoff (ed.), 127-138, New York: Academic Press.
- Liberman, M.Y. and L.A. Streeter (1978): "Use of nonsense-syllable mimicry in the study of prosodic phenomena", JASA 63, 231-233.
- Liberman, M.Y. and A. Prince (1977): "On stress and linguistic rhythm", Linguistic Inquiry 8, 249-336.
- Lindblom, B. (1963): "Spectrographic study of vowel reduction", JASA 35, 1773-1781.
- Lindblom, B., R. McAllister, and J. Lubker (1977): "Compensatory articulation and the modeling of normal speech production behavior", in Articulatory modeling and phonetics, R. Carré, R. Descout, and M. Wajskop (eds.), 148-161.
- Lovins, J.B., M.J. Macchi, and O. Fujimura (1979): "A demisyllable inventory for speech synthesis", in Speech communication papers, presented at the 97th Meeting of the Acoustical Society of America, J.J. Wolf and D.H. Klatt (eds.), 519-522.
- MacNeilage, P. (1970): "The motor control of serial ordering of speech", Psychol. Rev. 77, 182-196.
- Mattingly, I.G. (1977): "Syllable-based synthesis by rule", 9th International Congress on Acoustics, Madrid, July 4-9, 1977, Contributed papers 1, 512.
- McCutcheon, M.J., S.G. Fletcher, and A. Hasegawa (1977): "Video-scanning system for measurement of lip and jaw motion", JASA 61, 1051-1055.
- Mermelstein, P. (1973): "Articulatory model for the study of speech production", JASA 53, 1070-1082.
- Miller, J.E. and O. Fujimura (1979): "A graphic display for combined presentation of acoustic and articulatory information", in Speech communication papers, presented at the 97th Meeting of the Acoustical Society of America, J.J. Wolf and D.H. Klatt (eds.), 221-224.
- Minifie, F.D., C.A. Kelsey, J.A. Zagzebski, and T.W. King (1971): "Ultrasonic scans of the dorsal surface of the tongue", JASA 49, 1857-1860.
- Muller, E.M. and J.H. Abbs (1979): "Strain gauge transduction of lip and jaw motion in the midsagittal plane: refinement of a prototype system", JASA 65, 481-486.
- Nakajima, T. (1977): "Identification of dynamic articulatory model by acoustic analysis", in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.), 251-275, Tokyo: University of Tokyo Press.
- Nelson, W.L. (1979): "Automatic alignment of phonetic transcriptions of continuous speech utterances with corresponding speech-articulation data", in Speech communication papers, presented at the 97th Meeting of the Acoustical Society of America, J.J. Wolf and D.H. Klatt (eds.), 63-66.
- Perkell, J.S. (1977): "Articulatory modeling, phonetic features and speech production strategies", in Articulatory modeling and phonetics, R. Carré, R. Descout, and M. Wajskop (eds.).
- Pierrehumbert, J. (1979): "Intonation synthesis based on metrical grids", in Speech communication papers, presented at the 97th Meeting of the Acoustical Society of America, J.J. Wolf and D.H. Klatt (eds.), 523-526.
- Sawashima, M. (1979): "A supplementary report on speech production", Proc.Phon. 9, vol. I, 49-56.
- Sawashima, M. and H. Hirose (1968): "New laryngoscopic technique by use of fiber optics", JASA 43, 168-169.
- Shirai, K. and M. Honda (1977): "Estimation of articulatory motion", in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.), 279-302, Tokyo: University of Tokyo Press.
- Sondhi, M.M. and B. Gopinath (1972): "Determination of vocal-tract shape from impulse response at the lips", JASA 49, 1867-1873.
- Sonoda, Y. (1977): "A high sensitivity magnetometer for measuring the tongue point movements", in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.), 145-156, Tokyo: University of Tokyo Press.
- Sternberg, S., S. Monsell, R.L. Knoll, and C.E. Wright (1978): Information processing in motor control and learning, G.E. Stelmach (ed.), 117-152, Academic Press.
- Stevens, K.N. (1972): "The quantal nature of speech: evidence from articulatory-acoustic data", in Human communication, A unified view, P.B. Denes and E.E. David (eds.), 51-66, New York: McGraw-Hill.
- Titze, I.R. and D.T. Talkin (1979): "A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation", JASA 66, 60-74.

## DISCUSSION

H. Hirose, M. Hirano, and J.S. Perkell opened the discussion.

H. Hirose emphasized that we have to be careful in the interpretation of the electromyographic data, in particular because the relationship between the degree of muscle contraction and EMG output - in the case of speech muscles - is linear only under very special conditions. In order to get some idea of the relationship between the EMG pattern and the articulatory events we have to combine several methods. As an example, Hirose showed some EMG and X-ray microbeam data recorded simultaneously. He concluded that modern methods for the investigation of speech production can also be applied to the analysis of pathological patterns of movements and furthermore perhaps help towards a better understanding of the role of related parts of the central nervous system in speech production.

M. Hirano discussed various techniques employed for the study of the morphology and function of the vocal folds. He demonstrated that there are two different kinds of fibrous components in the vocal folds, namely the elastic and the collagenous fibres and showed how the vocal folds consist of more layers, partly from a histological and partly from a mechanical point of view (cf. vol. I, p. 189).

J.S. Perkell said that from his point of view the use of movement and EMG data along with sophisticated physiological modeling is only in its infancy with respect to the contribution that these techniques will eventually make to our understanding of speech production, dynamics, and hopefully also control strategies.

Then he commented upon the need for additional and better data in the third dimension, i.e. the cross-sectional area function, to supplement good midsagittal data. The need for such data is illustrated by the range of current notions that we have about factors, which underlie or constrain vowel categories. For example, B. Lindblom and his co-workers have proposed that a vowel category may be determined by an interaction between perceptual distance and some measure of ease of articulation; M. Lindau has proposed a primary role for the acoustic factors; K. Stevens has suggested some role for the patterns of tongue-to-maxilla contact; S. Wood and others have suggested that the vowel categories are determined

by quasi-discontinuous relationships between the place of constriction and the sensitivity of formants to changes in place and degree of this constriction, along with factors related to the muscular anatomy; and Fujimura, working with the tongue model, has suggested a role for discontinuous relationships between muscle contractions and area function. Perkell noted that we have no way of disproving any of these hypotheses, and it may well be the case that to some extent all of them are valid. He concluded that to begin to untangle all the possible influences on vowel categories, we need a lot more well controlled work to test each one of these hypotheses, and that improved knowledge of area functions along with other factors is obviously essential for the evaluation of all the hypotheses on articulatory correlates of sound categories.

Then Perkell turned to the question about the X-ray dosage for different X-ray techniques. Perkell and his co-workers have made some dosage measurements and he gave the following values for the dosage that the subject would get:

10 rads/min for 35 mm conventional cineradiographic film (60 frames/sec); 2,5 rads/min for 35 mm high speed film and for 60 mm conventional cineradiographic film; 600 mrad/min for 16 mm high speed cineradiographic film; 260 mrad/min for video-tape.

Perkell noted that the microbeam system rarely gets above approximately half of these values, but under most circumstances the microbeam exposes the subject to a much smaller dosage. Finally, Perkell mentioned that the X-ray unit they are using allows for simultaneous views in the anterior-posterior and in the lateral dimensions, and he hoped that they might be able to obtain information which will contribute to our insufficient knowledge of area functions.

O. Fujimura confirmed that the dosage for the X-ray microbeam system is about one half of the smallest dosage obtained with other X-ray systems, namely 120 mrad/min. But the frame rate used for the estimate of 120 mrad is 120 frames per sec., i.e. twice the rate that Perkell used for his estimate. And in order to derive the total energy absorbed into the body, the dosage given should be multiplied by the area under exposure; since the 120 mrad estimated for the microbeam system assumes a constant exposure over a small area of  $1 \text{ cm}^2$ , the product is obviously 120, whereas the exposed area is much larger in the case of the two other X-ray systems.

E. Keller found the ultra sound technique a valuable alternative to the X-ray technique. The great advantage is that the exposure time can be considerably longer than the exposure time using cineradiography. Keller also pointed out that the frame rate is limited with the X-ray methods, which is a problem if we want to make measurements of speed of articulation, for instance. Finally, Keller said that with the ultrasonic method using a scanning beam, i.e. a system where a beam is sent back and forth several thousand times per sec., the whole surface of the tongue can be recorded, for instance, contrary to what can be obtained by a single beam system.

O. Fujimura claimed that for the X-ray microbeam system the net total of exposure time for one session is typically about 10 min., and often they run two or three sessions per subject. The total dosage given to the subject in terms of energy absorbed in one session is comparable to the amount of dosage one gets from the cosmic rays during one year. Concerning the limitation of frame rate, Fujimura replied that if one is interested in studying very fast movements in one portion of the tongue, which is the normal application of the ultrasonic technique, the number of pellets can be reduced and thereby a frame rate of up to 1000 frames per sec. can be obtained, so the frame rate for the microbeam system is not restricted to anything like 120 frames/sec.

Finally, Fujimura mentioned that for the velum height measurements - using the X-ray microbeam system - the pellet is not glued directly on to the velum as is the case with tongue pellets. Instead, a narrow strip of a very flexible plastic sheet is inserted through the nostril, covering the pellet, and this keeps the pellet in position, in contact with the upper surface of the velum.

J. Ohala emphasized that the estimates of radiographic dosage that we find in the literature vary tremendously. Furthermore, he referred to a study revealing an increased incidence of cancer in the thyroid from a population who had been radiated 30 years ago as children, with a dose of 6 rads, but these cancers did not develop until now. Ohala concluded that though the vocal tract is very important for us, we have to be very cautious in estimating our dosages. He advocated an intensification of our search for alternative ways of getting vocal tract informations.

G. Fant mentioned the possibility of measuring the impedance between two points, e.g. the upper and lower lip, as an alternative method for tracking the dynamics of articulation.

H. Künzel mentioned a very simple instrument for real-time recording of velar elevation, developed at the Institute of Phonetics in Kiel. The system consists of an optical probe - with an outer diameter of 3 mm - inserted through the nostril. The probe emits light which is reflected as a function of velar elevation. The linear function of the system has been controlled by simultaneous X-ray recordings.

C. Scully mentioned another approach, which works back from the aerodynamic stage and infers movements of the articulators from aerodynamic data. Such a technique can give us some idea of the size of the constrictor across which a pressure drop can be measured. What sort of range and what degree of accuracy this yields is an open question at the moment, but it is being investigated.

O. Fujimura mentioned a new technique, suggested by Dr. Sinada, where the pellet position is detected purely magnetically. The only disadvantage is that at the moment only one pellet can be tracked.

The indirect methods are very useful, in particular for practical purposes like clinical applications, training of articulatory gestures, and so on. But they need calibration and here the microbeam system could also be used.

S. Smith claimed that the electroglottographic method tells us something about the state of the musculature, i.e. whether it is relaxed or contracted.

O. Fujimura said that a technique for measuring the state of the muscular contraction by some physical means would be advantageous if we can establish a way to calibrate it.

CORTICAL ACTIVITY IN LEFT AND RIGHT HEMISPHERE DURING LANGUAGE  
RELATED BRAIN FUNCTIONS

Niels A. Lassen and Bo Larsen,<sup>1</sup> Department of clinical physiology,  
Bispebjerg Hospital, DK-2400 Copenhagen, Denmark

Chairpersons: Peter Ladefoged and Hans Günther Tillmann

The blood flow through the brain cortex varies with the functional state of the tissue. Just as in skeletal muscle or in various glands, an enhanced level of nerve cell activity invokes an increase in tissue metabolism and in blood flow. Thus, it was found by Olesen from our group that rhythmical movement of the hand augments regional blood flow in the contralateral central (hand) cortex by 20 to 30 per cent. It was subsequently verified that indeed not only flow but also oxygen uptake is increased in that same area during hand exercise. We have used regional blood flow measurements to map the cortical areas active in various types of language related brain functions. A summary of our findings will be given.

The method used for measurement of regional cerebral blood flow  
in man

The radioactive isotope Xenon-133 is used. It is produced in a nuclear reactor as a split product of uranium. Like the non-radioactive Xenon isotopes, Xe-133 is an inert gas and (like nitrogen, N<sub>2</sub>) it does not react chemically with any molecules in the body. It is simply distributed according to the tissues' solubility. We use it in the form of a physical solution in saline in a dose of approximately 5 MilliCuries per injection (1.5 ml). The radiation exposure is negligible; it is much less than that of a single conventional X-ray study. This means that a series of repeated injections with an interval of 15 minutes can be made in the same setting without any radiation hazard. We take advantage of this by usually performing a series of 4 or 5 injections in one study: first at rest and then during a series of different forms of brain work - in this case involving various language related types of brain functions.

The Xenon-133 containing sterile saline is injected into a big artery on one side of the neck, the internal carotid artery. It supplies the anterior 3/4 of the brain (usually the posterior

---

1) The paper was given by N.A. Lassen.

part of the brain, the occipital lobe's inner side, is not receiving the isotope by this injection as its arterial supply comes from a different artery, namely from the vertebral artery). With each internal carotid supplying (normally) only the ipsilateral cerebral hemisphere, and by injecting only one side, we obtain maps of blood flow distribution in one hemisphere only. This is a distinct limitation with regard to studying hemispheric differences: we have to rely on comparing a series of left hemisphere observations with those on the right side in other subjects, and cannot in the same subject observe both sides simultaneously.

Using a special isotope camera with 254 small detectors, we observe the arrival and subsequent wash-out of the Xenon-133 in regions of the size of approximately  $1 \text{ cm}^2$ . The tissue element "seen" has the form of a cone traversing the injected hemisphere. Due to absorption of radiation it is, however, the superficial cortex we see best. The regional blood flow is calculated from the slope of the Xenon-133 wash-out curve during the first minute following the injection of the radioactive bolus (that takes only one second). When a test is performed, such as counting or reading, the subject is asked to start performing approximately 10 seconds before the Xenon-133 injection, and then continue for 60 seconds (the injection is not felt by the subject). The interval of approximately 15 minutes between injections is necessary in order to clear the brain of radioactivity before injecting the next dose (we can actually use a shorter interval, and then compensate for remaining radioactivity).

The technique is not entirely atraumatic: it involves the cannulation and injection into the blood flowing to the brain and a risk of compromising this flow exists. We have not encountered any complications in the series of 350 subjects studied in our laboratory (over a period of 4 years) with the technique described here. Yet, this risk restricts us to study patients with neurological symptoms in whom cerebral angiography is indicated, i.e. in whom a cannula is placed in the carotid artery for X-ray study. This means that normal subjects cannot be studied. Nevertheless, our series of neurological patients comprises cases without focal tissue abnormalities (patients studied because of arterial aneurisms or because of an epileptic seizure, cases of suspected brain tumor, etc.). The results obtained in such cases (approximately 20% of our patients) constitute our equivalent of normal

man. The main part of the studies reported below pertain to such "normal" cases. The consistency of the results leaves no doubt that the data may indeed be taken to pertain to normal man.

#### Results

A. The awake resting state. With closed eyes in a darkened silent room and completely at rest the normal pattern of blood flow distribution shows the highest values in the frontal lobe (approximately 10% above the hemispheric mean).

B. Listening to words. Simple noise produced with Barany noise apparatus increases flow in the hearing cortex only minimally. Listening to sounds (Seashore test) or onomatopoeica as "crack", "bang", "whiz", on the other hand, clearly activates this area on both sides (15-30% increase in flow). The area comprises Wernicke's center of language on the left side (all our subjects were right handed). Listening to music caused the same effect. Our data do not suggest a hemispheric difference with these two forms of simple listening tests.

Listening to more complex spoken language produces increased flow on the left side. But since this area overlies the basal ganglia and since a flow increase here is often seen with the unspecific more global flow increase accompanying increased attention, we cannot assert the specificity of this activation.

C. Talking. Automatic talk in the form of counting repeatedly to twenty at a rate of one digit per second activates the hearing cortex, the primary (rolandic) mouth area and the supplementary motor area.

All these changes are bilateral. The pattern tends to be less sharply demarcated on the right side than on the left.

Word naming in the form of finding words of 5 flowers, 5 types of furniture, etc., activated the same three areas and caused a constant activation of the whole prefrontal region as well (cf. the comments made under reading aloud and internal speech).

D. Reading aloud. This activates six areas in both hemispheres. In addition to the three areas seen during automatic talk, the following areas also become active: the visual association cortex in the posterior part of the brain, the frontal eye field that often merges with the mouth area, and the low-posterior part of the frontal lobe (with Broca's area on the left side) which we commented on above.

We cannot - in most of our cases - see the primary visual cortex as it is usually supplied by the vertebral artery. But from animal studies it is evident that this area becomes more active during visual stimulation. Hence, including this area, a total of fourteen discrete cortical areas, seven on each side, are active during reading aloud. Often, the prefrontal cortex anterior to the supplementary motor area, is also activated.

Reading a text aloud is a prime example of the fundamental mode of operation of the cerebral cortex in performing complex tasks (and there are probably no simple ones!): collaboration between discrete cortical areas, each performing a specific job. It is the pattern of activation that is related to the task, not any single area. There are, in other words, no isolated center solely responsible for solving a complex task as also emphasized by the late Alexander R. Luria.

So far we have not been able to discern individual patterns of cortical activity of such a nature as to suggest fundamental differences between individuals.

The role of the right hemisphere in this complex language function is not clear. But data from the literature suggest that production and analysis of language melody and perhaps even of gestures related to speaking may predominantly reside on the right side.

E. Reading silently. If the same subject after reading aloud reads silently, the change in the map of blood flow, compared to that at rest, is particularly easy to interpret: then the primary sensori-motor mouth area and the auditory cortex do not become active. All the other areas are, however, seen to be active.

F. Internal speech. Memorizing the text internally causes a small increase in the mean blood flow, predominantly in the frontal lobe (often especially in the prefrontal cortex).

This type of "global" activation is seen with any task that the subject makes an intellectual effort in accomplishing.

In our opinion, internal speech is a mental function which is just as real as love, hate, or memories. It is a solid fact of introspection. This is supported by the fact that one can readily think in different languages. But, while asserting the psychological reality of internal language, we would consider it imprudent to follow Luria's speculation that this language function has a special grammatical construction.

It is tempting to revert the argument and to state that internal speech is the only true or "essential" language function.

How about a patient paralyzed by a disease or Curare and who tries to speak? Can one state that he has no language function? In a way, all the external manifestations of language functions are non-essential - solely the internal functions of language understanding and production - both comprized in the concept of internal speech - are truly essential.

G. Aphasia. We have studied the blood flow map on the left side in a series of classical aphasia patients, mostly cerebro-vascular accidents ("apoplexy"). Confirming well-known facts, the "fluent" aphasia cases had defects (on the flow map) in the posterior speech area of Wernicke, "non-fluent" or "motor" aphasia had defects in the primary mouth area (sometimes but not always extending to Broca's area), "global" aphasia had large defects covering both Wernicke's area and the mouth area. No studies were made on the right hemisphere.

H. Auditory agnosia (comprising word deafness). A rare case of sensory aphasia due to bilateral temporal lobe infarcts was studied in some detail. The patient, a 63 year old man, first suffered an attack of mild fluent aphasia, lasting one month. Some months later, he suddenly lost all ability to understand any spoken words and had some difficulty in recognizing non-verbal sounds. Yet, his threshold for perceiving pure tones was normal for his age. In other words, he was not deaf. But he could not identify any words. Not even his own name, or simple words, such as yes and no. All other language functions were intact: talking, reading, writing. This state is in neurological terminology called "auditory agnosia".

Specialized investigations suggested that an acute right-sided lesion of the hearing cortex had cut off ("disconnected") the remaining posterior part of the left superior temporal gyrus (Wernicke's center) from its remaining input (that from the left side having been destroyed by the first stroke). Computerized tomography (CT-scanning) showed the bi-temporal infarcts as hypodense lesions involving Heschl's gyrus bilaterally. Regional blood flow studies during listening to sounds showed no activation of the upper part of the left temporal lobe area (Wernicke's area): The sound analyzer was not turned on!

This case is interesting for three reasons:

- 1) A right hemisphere lesion (lesion no. 2) produced a massive language handling defect in a right-handed subject. There are other cases of this type recorded in the literature.

2) Preservation of normal speech in a subject deprived of all meaningful auditory feed-back. The fact that completely deaf patients can speak fairly normally also confirms that some of the speculations concerning the necessity of the normal auditory feed-back for speech production have been exaggerated. The importance of this feed-back for language acquisition is not questioned in this argumentation.

3) The patient had completely normal early components in his auditory evoked response. Hence this response cannot originate in the primary auditory cortex -Heschl's gyri- as these were massively destroyed bilaterally.

#### Concluding comments

Many linguistic and phonetic problems related to cortical function could be posed in relation to the findings we have summarized in this paper. However, it is appropriate here to stress the poor temporal resolution (1 minute) and spatial resolution (1 cm<sup>2</sup> with superposition of deeper layers) involved in the registration of the regional blood flow. Certainly, we cannot by this approach say anything about the detailed way in which the cortical areas collaborate in language functions.

It surprised us to find that a simple sound-rhythm discrimination test (Seashore) activated the auditory association cortex to much the same extent as do music or language. Apparently, the whole sound analyzer works as a unit.

The major finding was in our opinion the bilateral and practically symmetrical cortical involvement in all language functions. The possibility of a special role of the right side for prosody is mentioned. We have no data pertaining specifically to this point.

A comment on memory may be appropriate. It appears that this function is disseminated in the brain: visual memory in the visual association cortex, tactile memory in the sensory cortex, etc. Thus it is not surprising that word memory resides in the auditory association cortex in the temporal lobe. That it is predominantly on the left side is, however, completely mysterious! Could it be that the speed of language perception (and production) precludes major inter-hemispheric information exchange in this most human or "highest" of all types of brain work?

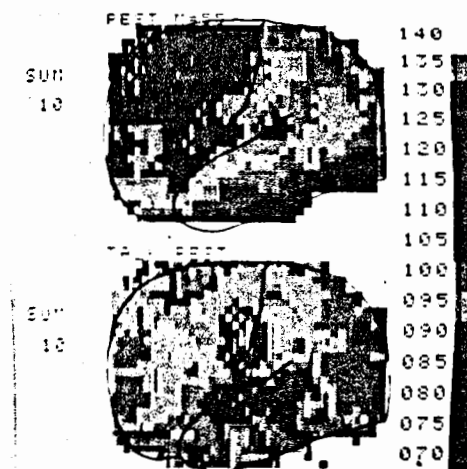


Fig. 1 Intact normal man, regional cerebral blood flow, rCBF, map, left hemisphere.

The original illustration is in colours and therefore the black and white reproduction has distorted the scale. The legend to this figure gives a verbal description of the increase in flow clearly seen on the original and also visible on this reproduction.

The upper frame shows the rCBF map at rest, average picture of 10 cases. The map is expressed in percent flow deviation from the mean hemispheric value (averaging 55 ml/100g/min in these cases).

The lower frame shows the average rCBF map during automatic speech expressed as percentage deviation from the map at rest. Three areas of consistent flow increase ("activation") are seen (in this black and white reproduction the areas are slightly darker than the rest, with still darker edges): the supplementary motor area (at the top), the sensory-motor mouth area (upper mid), and the posterior part of the superior temporal gyrus (lower mid, Heschl's gyri and Wernicke's area).

Changes in the right rCBF map during automatic speech are practically the same. Broca's area is usually seen with fluent speech.



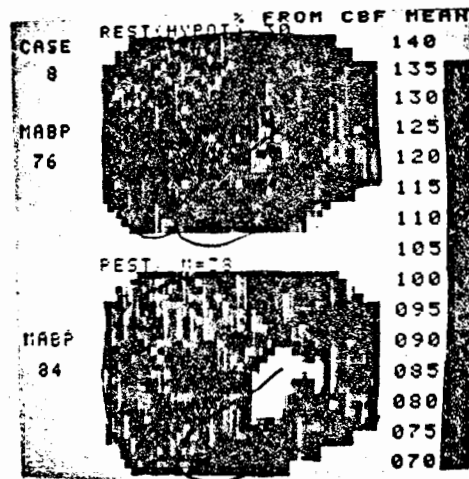


Fig. 2 Stroke case with aphasia (case 8 of our series), regional cerebral blood flow map, left hemisphere.

The upper frame shows rCBF map at rest during normotension (mean arterial blood pressure 84, mean rCBF 38 ml/100g/min).

Note the dramatic increase of flow in Wernicke's area (white) as flow rises: Luxury perfusion 8 days after onset of stroke, probably overlying an infarct, with abnormal pressure passive flow regulation.

#### References

- Larsen, B., E. Skinhøj, and N.A. Lassen (1978): "Variations in regional cortical blood flow in the right and left hemispheres during automatic speech", *Brain* 101, part 2, 193-209.
- Lassen, N.A., D.H. Ingvar, and E. Skinhøj (1978): "Brain function and blood flow. Changes in the amount of blood flowing in areas of the human cerebral cortex, reflecting changes in the activity of those areas, are graphically revealed with the aid of radioactive isotopes", *Scientific American* 239, 62-71.
- Soh, K., B. Larsen, E. Skinhøj, and N.A. Lassen (1978): "Regional cerebral blood flow in aphasia", *Arch. Neurology* 35, 625-632.
- Lassen, N.A. (1978): "Cerebral blood flow in cerebral ischemia. A review", *European Neurology* 17, suppl. 1, 4-8.

#### DISCUSSION

Victoria Fromkin, Michael Studdert-Kennedy and Peter MacNeilage opened the discussion.

Victoria Fromkin quoted Fournier's statement (in the late 19th century): "Speech is the only window through which the physiologist can view the cerebral life" and added that it should also be recognized that the brain is a window through which we will be able to observe the linguistic life.

Victoria Fromkin then expressed her hope that these new techniques would reveal to what degree language is a special function of the brain rather than a particular case of more general faculties. We ought to find out whether patients show differences in brain activity when they are subjected to stimuli of varying degrees of phonetic or linguistic complexity. And she mentioned that there might be different reactions to known versus unknown language stimuli, which again might be different from clearly non-linguistic input. Finally, different reactions might also be expected from patients automatically repeating memorized formulae rather than producing or reacting to free, creative speech.

In connection with the supposedly unexpected activity of the right hemisphere during automatic speech Victoria Fromkin mentioned that it is well known that even people who display a marked hemispheric specialization will always show some activity even in their right hemisphere during speech.

Victoria Fromkin further pointed to the dangers of drawing too far-reaching conclusions from observations based solely on patients with abnormalities of the brain. And she stressed the importance of looking for a convergence of results from different techniques.

Finally, she mentioned the importance of sensory aphasia cases, such as had been described in the lecture, for the debate on whether grammar exists apart from perception and production.

Michael Studdert-Kennedy: I think it is quite clear that techniques of this kind, such as the blood flow techniques, the more advanced analyses through EEG work, and perhaps the development of cooling techniques for isolating parts of the brain in the normal brain, are going to be much more important in the future than the type of behavioural studies that we have had to rely on in the past.



Michael Studdert-Kennedy then mentioned the problem raised by Victoria Fromkin in that these techniques are always used with patients. And he pointed out that in cases of apoplexy the right hemisphere could slowly be taking over functions normally performed by the left hemisphere. Therefore these new techniques should be developed and be made usable with normals.

He then continued: Obviously, the finding that there is a large amount of right hemisphere activity as well as left hemisphere activity is not a surprise. Because presumably there is a coordination of function between the two sides of the brain.

Nonetheless, there are certain properties of one side of the brain rather than the other that do arouse interest. And that seems to me to be important in understanding the nature of linguistic communication. I am referring here particularly to the famous relationship between speech and handedness.

It seems to me that an understanding of that relationship would take us rather a long way to understanding what the prior signalling conditions are for communication.

In this regard I think that the current developments in the work on sign language is tremendously important. Because it does seem that a prerequisite for linguistic communication is a motor system that is capable of very fine, rapid articulation.

One has only got to ask oneself what sort of a sign language could be developed if one was forced to use one's feet to realize that one absolutely has to have pieces of machinery that can be moved very, very fast. And so the motor control of that machinery which appears to be in some way common between the speech mechanism and the hand mechanisms, are of great interest. And I think that one very exciting possibility that these techniques look forward to is an elaboration and an understanding of these links.

In that respect I wonder, too, what the prospects are for looking at these processes developmentally.

Michael Studdert-Kennedy then drew attention to the sensory-motor integration functions described by Niels Lassen, particularly those concerned with speech and hand movements. These integration functions would appear to be a necessary prerequisite for the development of language. And he mentioned how children exposed to sign language will start imitating this at the same time as spoken language is normally developed.

Finally, he, too, suggested that the new techniques be used with different types of linguistic stimuli.

Peter MacNeilage was particularly interested in the spreading of activity from the temporal to the parietal lobes, having observed in the slides an upward spreading of activity in the parietal lobe during reading as compared to counting, and a still further spreading during listening. And he continued: The reason I am interested in the parietal lobe is its involvement in what is usually called conduction aphasia, and because I believe at the moment that the posterior and inferior parietal lobe is of some importance in the formulation of complex, voluntary movements.

Peter MacNeilage, too, mentioned the possibility of reorganization of brain functions after hemispheric lesions. And he suggested in the specific case mentioned by Lassen that possible simultaneous damage to Heschl's gyrus should be considered.

He then said: You may not exactly have intended to say this. But when you were talking about the finger movement task, you pointed to the fact that there was a rather circumscribed and small area of high activity in area four, that did not extend very anteriorly. On the other hand, there was a much larger and more widespread area of activity in the somatic-sensory cortex. And I believe you said that you thought the somatic-sensory activity was of more importance than the motor activity. I would like you to clarify this remark. [Niels Lassen: "That is correct."] Because it seems to me to relate to a rather general question about the extent to which we can simply assume a linear relation between the amount of activity, or wideness of distribution of activity, and the importance of the function.

It seems from my point of view that there may be parts of the cortex that can get their job done with less blood flow than other parts.

In your Scientific American paper you talked a little more of the role of the supplementary motor cortex than you have here. You still believe that the supplementary motor cortex has an important organising role in the production of speech? Because an alternative hypothesis is possible, namely that it simply has to do with initiation, or facilitation, of action in a rather general sense.

One could possibly argue that it has the equivalent of an attentional role on the motor side. It facilitates things happening without actually having much to do with the details of the control function themselves.

Coming back to the question of skilled, voluntary movement, I would like to ask to what extent you have studied unskilled voluntary control versus skilled voluntary control. That is in particular in relation to learning a skilled voluntary task.

And finally, I have heard that Brenda Milner, using sodium amytal studies, has shown a rather interesting relationship between the controlling hemisphere in left handers for speech, and for skilled voluntary movement of other kinds.

Niels Lassen, answering the first three discussants, said how surprised he and his colleagues had been when they saw to what degree the entire auditory association cortex was activated when a patient was stimulated with even very simple sounds. Stimulation with longer sequences and more complex stimuli was found to raise the level of activity a little more, but in the same area. Since the difference in reaction to simple and complex stimuli was found to be so small, the general rise in activity may be thought of as a sort of local attention phenomenon, where the whole auditory system is activated by any incoming signal.

Concerning the proposals to use more differentiated stimuli, Niels Lassen mentioned that he had received a stimulus tape from the Phonetics Institute, Copenhagen, containing white noise, isolated vowels, simple CV-syllables as well as connected speech, and was planning to use this in further experiments.

About the questions concerning the supplementary motor area, Niels Lassen said that he had found this area particularly active during complex movements. The relatively high level of activity in this area during speech could therefore be explained by the fact that speech is produced by very fast and complex movements.

As to the question about Heschl's gyrus, Niels Lassen said that there were damages to that area on both sides, but that he was not sure whether it had been completely destroyed. What was clear was that there no longer arrived any information to the auditory association cortex on the left side. That was evident from the lack of flow increase in that area when listening to words.

Niels Lassen confirmed that the parietal lobe is indeed very active, also during speech. But this appears to be part of the general arousal of the brain, since this area becomes active with any kind of activity on the part of the patient.

Barbara Prohovnik mentioned that people in Lund were using Xenon inhalation methods, which are non-invasive, to obtain similar traces.

Niels Lassen answered that the inhalation methods gave less well defined results because of limitations in the time constants of these methods.

Prompted by several people, Niels Lassen stated that the method he described measures the average activity of the brain over at least ten or fifteen seconds. Generally, a recording averages over the first thirty or forty seconds where the important information is concentrated. It is known from animal studies that there is a time lag of two or three seconds from the time of the injection to the time when changes in the blood flow can be clearly detected, and it disappears over ten to fifteen seconds. A new injection may be made after about three minutes. But successive recordings have even been made with only one minute intervals.

Vincent van Heuven suggested that we look not only for the areas of increased blood flow, but that we also examine what areas are inactive during a particular task. Both Vincent van Heuven and Niels Lassen commented on the fact that the brain always shows activity somewhere, even when the patient is at rest. But Niels Lassen said that he had observed not only increases but even reductions in the level of activity in certain areas when the level rose in other areas because the patient concentrated on a particular task.

John Laver was sceptical about the reported case of a bilateral lesion which had made auditory feed-back impossible. Experience shows that this should cause a progressive deterioration of the articulatory accuracy of the patient's speech, which apparently it had failed to do in this case.

H. Mol mentioned that he knew of a totally blind and deaf man, who has excellent speech performance. His deafness developed suddenly at the age of 31 as the result of meningitis.

Niels Lassen said that this case, just as his own, strongly supports the notion of the unimportance of auditory feed-back for speaking a well established language.

## NEW METHODS OF ANALYSIS IN SPEECH ACOUSTICS

Hisashi Wakita, Speech Communications Research Laboratory Inc.,  
806 West Adams Boulevard, Los Angeles, California 90007, U.S.A.

Chairperson: Hans Werner Strube

Introduction

The recent development in digital techniques has brought substantial innovations to methods and techniques for acoustical analysis of speech sounds. The advantages of using digital computers over the conventional analog techniques are that the analysis processes can be repeated precisely and that the control of the parameters is relatively easy. The use of the digital computer also permits the processing of a large amount of data within a relatively short period of time with satisfactory accuracy. Because of the above advantages, digital techniques are playing a more and more important role in speech research. As this tendency becomes stronger, proper care has to be taken when the digital techniques are applied to speech research. This paper, thus, concerns primarily the recent digital techniques in the acoustic analysis of speech, particularly the linear prediction method, with special attention to its advantages and disadvantages, and also to the limitations involved in the technique.

The concept of linear prediction was first applied to speech analysis by Itakura and Saito in Japan (1966) and by Atal and Schroeder in the United States (1967). Since then the linear prediction method has been fairly thoroughly studied theoretically and experimentally (see Makhoul 1975; Markel and Gray 1976; Wakita 1976), and the method is currently being used as a powerful tool for acoustical analysis of speech sounds.

Linear prediction of speech

A very simplistic model of speech production as shown in Figure 1 (a) is assumed in the linear prediction of speech. The excitation source is an impulse and the filter, which mainly represents the vocal tract, has the frequency characteristics of resonances only, without any anti-resonances. The model thus exclusively represents the voiced and non-nasalized sounds.

For an analysis model, an inverse filter is assumed, which maintains the precise inverse relation between the input and the output of the production model, as shown in Figure 1 (b). Thus,

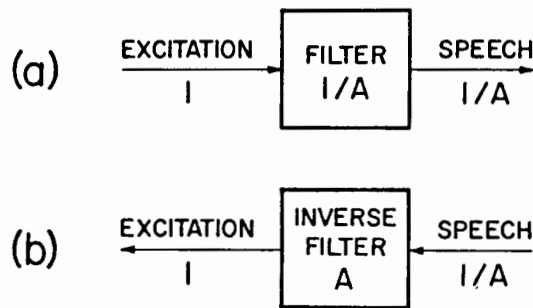


Figure 1. Models for the linear prediction method: (a) Production model; (b) Analysis model.

the problem in linear prediction analysis is to determine the characteristics of the inverse filter from a given input speech wave.

Since the linear prediction method is a digital technique, all the data, and parameters to specify the filter characteristics, are handled in a discrete sampled format instead of as continuous quantities. The main task of linear prediction is to predict the current speech sample  $\hat{x}_n$  in terms of a linear combination of the past M samples. Letting the predicted current sample be  $\hat{x}_n$ ,  $\hat{x}_n$  is given by

$$\hat{x}_n = \alpha_1 x_{n-1} + \alpha_2 x_{n-2} + \dots + \alpha_M x_{n-M} \quad (1)$$

In equation (1), the  $\alpha_i$ 's are called predictor coefficients. They play a role of "weighting" the past samples to predict the current one. The problem in the linear prediction method is to determine these predictor coefficients in such a way so as to minimize the error between the current sample and the predicted one, and to relate the predictor coefficients to the parameters of the inverse filter. In this case, the sum of the squared errors over a certain period,

$$E = \sum_{n=1}^N (x_n - \hat{x}_n)^2 \quad (2)$$

is minimized. Because of this, speech samples during this period are assumed to be sufficiently stationary so that the predictor coefficients do not change during this period.

How are the predictor coefficients thus determined related to physically meaningful parameters, that is, to the inverse filter in Figure 1 (b)? In general, the frequency characteristics of a filter can be determined by observing its impulse response when an impulse signal is applied to the filter as shown in Figure 2 (a). In the discrete case, the impulse response of a filter is then given as shown in Figure 2 (b). The amplitude at each sampled point in the impulse response is given by  $a_i$  and the period between the two sample points is given by the sampling period T. From this impulse response, the transfer function, A(z), of the filter is given by use of "z-transform" notation as

$$A(z) = a_0 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_M z^{-M} \quad (3)$$

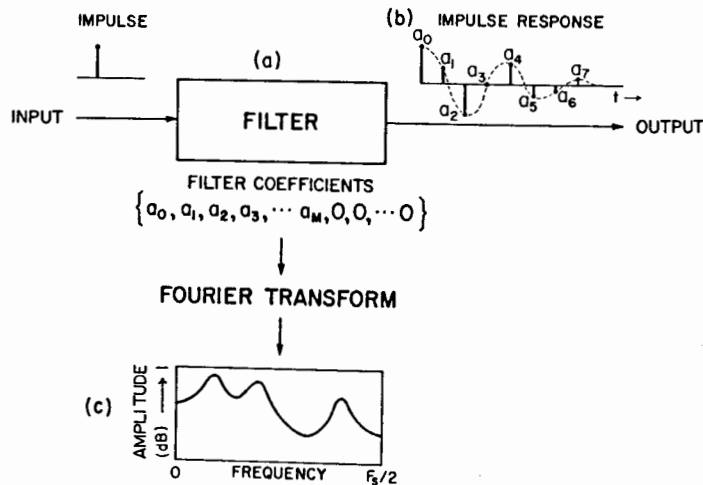


Figure 2. Determination of filter characteristics: (a) a model; (b) discrete impulse response; (c) frequency characteristics (transfer function) of the filter.

Equation (3) represents not only the transfer function of the filter but also the impulse response in the time domain. The  $a_1$ 's in equation (3) are called filter coefficients. It is easily seen from Figure 2 (b) that the interpretation of the "z-transform" notation is that  $z^{-1}$  represents a unit delay in the time domain in terms of the sampling period  $T$ . Thus, the power of  $z^{-1}$  in equation (3) denotes the number of time delays.

Since  $z = \exp(j2\pi fT)$ , where  $j$  is the imaginary unit ( $j = \sqrt{-1}$ ) and  $f$  is frequency, equation (3) itself represents the discrete Fourier transform of the impulse response. Thus the frequency domain representation of equation (3) is given by applying the Fourier transform to the filter coefficients. In this case, the impulse response is truncated at  $t = MT$  and normally sufficient zeroes (e.g. 256 minus  $M$  zeroes) are added to the  $a_1$ 's to ensure sufficient frequency resolution before the Fourier transform is applied. An example of a power spectrum obtained from the output of the Fourier transform is given in Figure 2 (c). Note that the frequency band is bounded at  $F_s/2$  where  $F_s = 1/T$  is the sampling frequency. Note also that when the amplitude of the frequency components is represented on a logarithmic scale, the frequency characteristics of the inverse filter as shown in Figure 2 (c) become those of the vocal tract filter in Figure 1 (a) just by re-labeling the negative sign of the ordinate with a positive sign.

One of the important features of the linear prediction method is that the predictor coefficients in linear prediction of speech can be shown to be identical to the filter coefficients with  $a_0 = 1$ . Consequently, minimizing the overall error in linear prediction is equivalent to finding the transfer function of the inverse filter of the analysis model in Figure 1 (b).

#### Analysis condition

Proper analysis conditions for the linear prediction method are important to ensure satisfactory results. The analysis conditions to be noted are (1) sampling frequency, (2) the number of coefficients, (3) time window and length, (4) window shift, and (5) preemphasis. The sampling frequency determines the frequency range of interest. The frequency range must be less than or equal to half the sampling frequency (normally the latter is chosen). The number of coefficients is dependent on the frequency range to be chosen. When the frequency range is exactly half the sampling

frequency ( $F_s$  kHz), a good rule of thumb for the number of filter coefficients is from  $F_s + 2$  to  $F_s + 4$ . The reason for this appears to be that there will be about  $F_s/2$  resonances in the frequency band limited by  $F_s/2$ , provided that  $F_s$  is given in units of 1 kHz. Each resonance requires 2 coefficients for its representation, and so about  $F_s$  coefficients will be needed to account for the expected resonances in the analysis band. In addition, 2-4 coefficients are normally used for approximating the spectral slope due to the excitation source.

The analysis conditions (3) and (4) vary depending upon which of two different methods of linear prediction is used, the autocorrelation method or covariance method (e.g. Markel and Gray 1976). The two methods use different definitions for computing the coefficients from sampled speech. The autocorrelation method requires a window length of at least 1.5 pitch periods and a Hamming window is recommended to suppress the spectral disturbances in the high frequency region due to the edge effect of the time window. The covariance method, on the other hand, does not require any particular time window, and the window length can be less than a pitch period. Thus this method can be used for pitch-synchronous analysis of speech sounds. When a window length of less than a pitch period is chosen, care must be taken since the analysis results vary depending upon what portion of the pitch period is chosen for analysis. This method is particularly useful for extracting the true vocal tract characteristics by choosing the glottis-closed portion of the speech waves. The major disadvantage of the covariance method is that there is theoretically no guarantee for obtaining a stable transfer function for the inverse filter, and thus a more sophisticated algorithm is required to automatically process the cases of instability. Also a more sophisticated algorithm is needed for automatically windowing the speech wave into pitch-synchronous intervals.

The window shift in the covariance method, thus, involves a more complicated procedure than it does in the autocorrelation method. In the latter method, the window shift is rather arbitrary, depending upon the speech samples to be analyzed. The shift can be greater than the window length for steady-state sounds, whereas, for speech sounds in which the formant frequencies are rapidly changing, a smaller window shift will be better for obtaining the smooth contour of the formant frequencies.

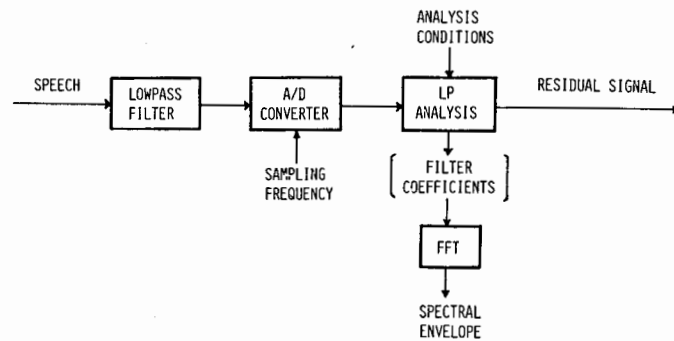


Figure 3. A block diagram to compute the smooth spectral envelopes of speech sounds by the linear prediction method.

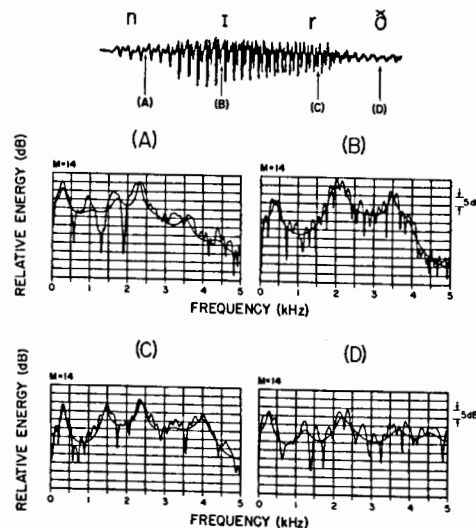


Figure 4. An example of linear prediction analysis. (Sampling frequency 10 kHz; number of coefficients 14; window size 20ms with a Hamming window and +6dB/octave preemphasis.)

A 6 dB/octave preemphasis is recommended for formant analysis. This is accomplished by taking the backward differencing of the sampled speech. The purpose of the preemphasis is to enhance the spectral peaks in the high frequency region. The 6 dB/octave preemphasis also roughly compensates the -12 dB/octave glottal source characteristics and the +6 dB/octave lip radiation characteristics.

Estimation of formant frequencies

As mentioned before, the Fourier transform of the predictor coefficients gives the frequency characteristics of the inverse filter, the inverse of which are the frequency characteristics of the vocal tract filter. Thus the procedure for obtaining the smooth spectral envelope by use of the linear prediction method is given by the block diagram shown in Figure 3. The speech signal is first digitized at some sampling frequency after being passed through a lowpass filter to limit the frequency band according to the sampling frequency. Linear prediction analysis is then performed using predetermined analysis conditions, and resulting in a set of filter coefficients for each speech segment analyzed. Smooth spectral envelopes are computed from the output of the Fourier transform of the filter coefficients with added zeroes. As a result of linear prediction analysis, the residual signal, which is an error signal given by equation (2) is saved for detecting pitch periods as will be described later.

An example of analysis results is shown in Figure 4. This example is a part of a sentence "Near the boat ..." and the spectral envelope estimation for /n/, /l/, /r/, and /ø/ are shown in the figure together with the direct Fourier transform of the corresponding speech waves. It is seen that spectral peaks are well approximated by the extracted spectral envelope. However, the spectral dips due to anti-resonances as in the sound /n/ are ignored in the linear prediction method, in which the nasal tract is not considered. It should be noted that the linear prediction method was developed as a method for efficient speech analysis-synthesis telephony on the basis of the fact that the human ear is insensitive to spectral dips. Thus ignorance of spectral dips is not a major problem as far as analysis-synthesis telephony is concerned. However, if one is interested in more accurate estimation of spectral dips as well as peaks, a new model has to be developed, which is currently being investigated by some researchers.

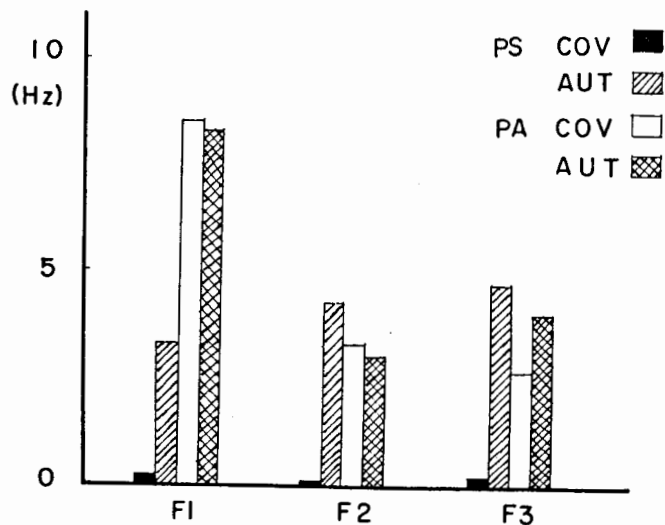


Figure 5. Evaluation of formant frequency estimation by autocorrelation and covariance methods for pitch-synchronous and pitch-asynchronous cases.

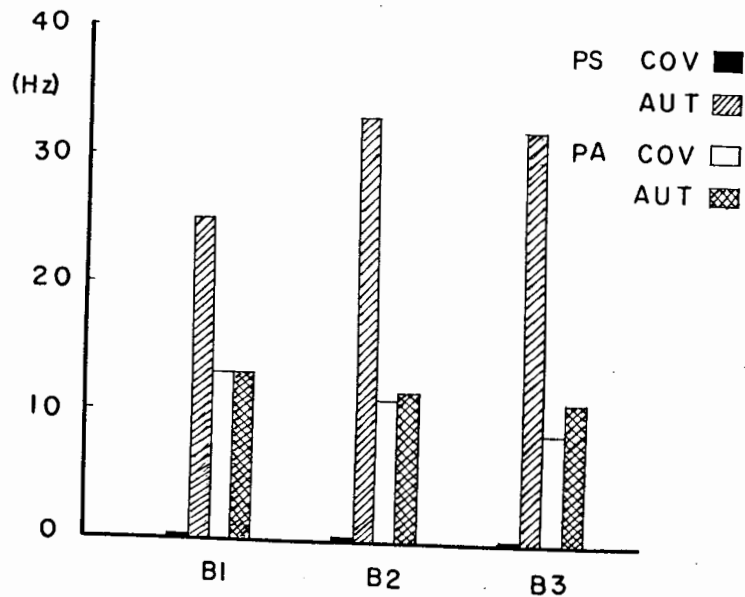


Figure 6. Evaluation of formant bandwidth estimation by autocorrelation and covariance methods for pitch-synchronous and pitch-asynchronous cases.

The formant frequencies are estimated from the smooth spectral envelope by finding the locations of the spectral peaks by a peak-picking method. Although this method is simple and worthwhile, it presents problems when two peaks are close together or merged into a broad peak. Another method is to compute the exact locations of the peaks by solving for the roots of the transfer function,  $A(z)$ , of the inverse filter. In both methods, the spectral peaks do not always correspond to the formant frequencies, and thus a certain algorithm to automatically select formant peaks has to be designed (e.g. McCandless 1974). For both methods, a careful inspection of the analysis results is recommended before further processing of the formant frequencies is initiated.

Accuracy of formant estimation

It is rather difficult to determine the accuracy of formant estimation for natural utterances, since there is no way of accurately measuring the vocal tract configuration to compute its resonances while a sound is being produced. Chandra and Lin (1974) made an evaluation of the autocorrelation and covariance methods of linear prediction by using synthetic vowels. In their study, vowels in the 'h-d' context were synthesized by a simulated formant synthesizer, and the two linear prediction methods were applied to analyze those synthetic vowels. As analysis conditions in this case, the sampling frequency was 10 kHz and the number of coefficients was 12. The results of their study are shown in Figures 5 and 6. Figure 5 shows the estimation error (in Hz) of the first three formant frequencies for both methods applied pitch-synchronously and pitch-asynchronously. For the pitch-synchronous case, the window length coincided with the segment position between the two pitch pulses. For the pitch-asynchronous case, the window length of 24 ms was arbitrarily chosen on the speech waves. The results indicate that the pitch-synchronous covariance method gives better accuracy than the others. In the pitch-asynchronous case, when the window length becomes greater than one and a half pitch period, the two methods give similar accuracy. The pitch-synchronous autocorrelation method resulted in the worst accuracy. This is more so in estimating formant bandwidths as shown in Figure 6.

For natural utterances, it is anticipated that the accuracy of estimating formant frequencies and bandwidths becomes worse

than for the synthetic sounds. Especially, it is anticipated that the result of the pitch-synchronous case will become worse, because the condition at the glottis varies during one pitch period for natural utterances, whereas the glottal condition for this particular synthesizer was constant. When the glottal condition varies during a chosen analysis segment, the resulting formant frequencies will probably be the average of the instantaneous formant frequencies. The result obtained by Chandra and Lin (1974) indicate that the pitch-synchronous covariance method gives more accurate estimates of formant frequencies and their bandwidths than the pitch-asynchronous autocorrelation method. Although the estimation accuracy of the formant bandwidths is not well known, it is known that the bandwidth estimates are sometimes too narrow or too broad. If the bandwidth information is needed, it has to be carefully checked against the direct Fourier transform of the corresponding sampled speech.

#### Problems in formant estimation

Since the estimation of formant frequencies is made from the envelope estimation of speech spectra, the accuracy of estimation is highly dependent on harmonic density. The more sparse the harmonic density becomes as pitch goes up, the more difficult the estimation of formant frequencies becomes. This is a rather inherent problem in the estimation of vocal tract resonances from given speech waves, irrespective of method. In many cases, the linear prediction method works well for speech sounds with fundamental frequencies of up to approximately 250 Hz. For female speakers and children with fundamental frequencies higher than 250 Hz, difficult cases of formant estimation are frequently observed. Formant estimation becomes impossible as the pitch becomes extremely high, in which case harmonics are picked up as spectral peaks.

In case the exact vocal tract resonances need to be known, some other methods may have to be used. One approach to this is to use external excitation with a low fundamental frequency such as an artificial larynx buzzer. One such example is shown in Figure 7 (a). This example is a female vowel /a/ with a fundamental frequency of 250 Hz. The linear prediction spectral envelope has one broad peak in the low frequency region instead of the first two formant frequencies. The peak-picking method de-

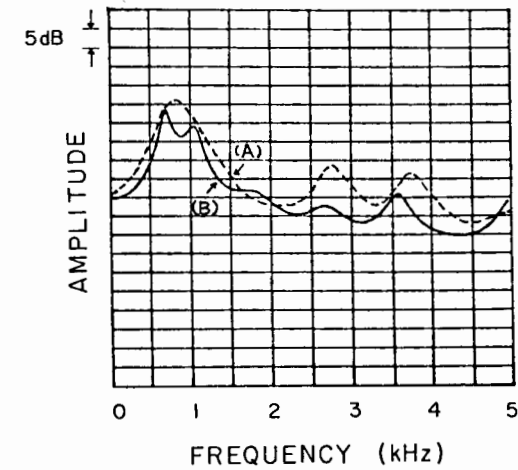


Figure 7. An example of difficult case of formant estimation. (a) Linear prediction spectral envelope for the vowel /a/ by a female speaker with a fundamental frequency of 250 Hz (sampling frequency 10kHz; number of coefficients 12; window size 25.6ms with a Hamming window and +6dB/octave preemphasis). (b) Linear prediction spectral envelope for the vowel /a/ by the same speaker excited by an external buzzer with a fundamental frequency of 80Hz (analysis conditions are the same as in (a)).

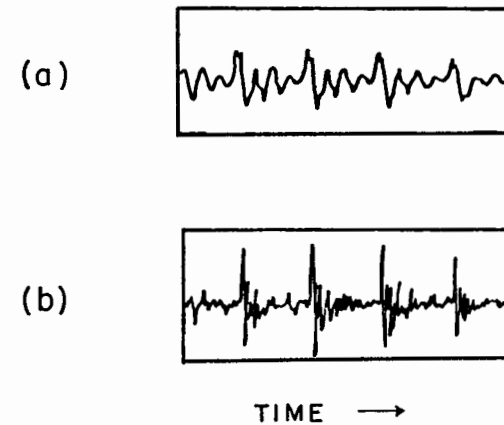


Figure 8. (a) Speech waves; (b) the residual signal after linear prediction analysis.



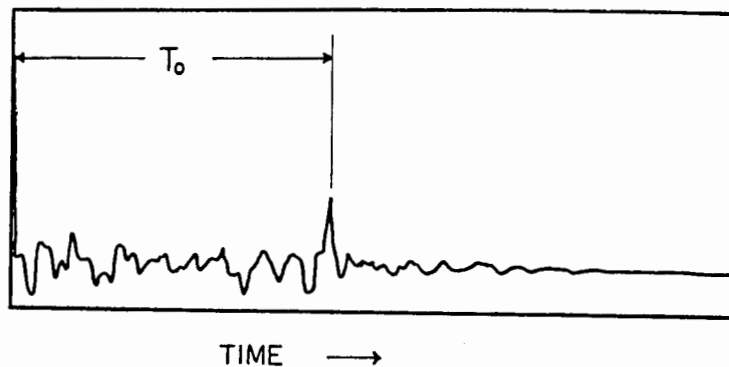


Figure 9. Autocorrelation function of the residual signal in Figure 8.

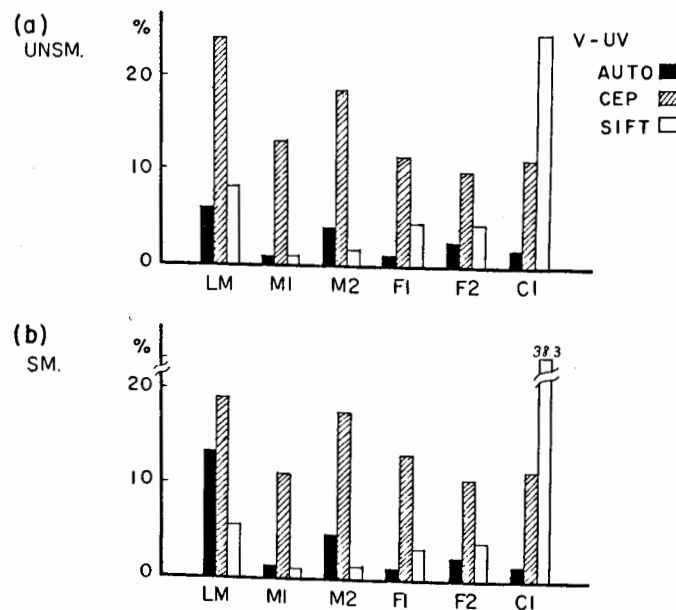


Figure 10. Voiced-to-unvoiced errors for three pitch detection methods: (a) unsmoothed; (b) smoothed. (LM: low-pitched male; M1, M2: males; F1, F2: females; CI: child). The ordinate shows the percentage error rate against total number of voiced intervals.

finitely fails to detect two peaks for  $F_1$  and  $F_2$ . Instead it will detect the broad peak as the first formant frequency.

The root-solving method will give two roots to approximate the broad peak. It has not been ascertained, however, that the two roots obtained by the root-solving method for such cases as above correspond accurately to the first two formant frequencies. For the above case, the use of a commercial artificial larynx buzzer with a low fundamental frequency gives a good resolution for the formant frequencies as shown in Figure 7 (b), which is for the same vowel and the same speaker as in Figure 7 (a). In this case, the buzzer had undesirable sharp peaks in its own frequency characteristics. The monotonous frequency characteristics of a buzzer are desirable for this purpose.

#### Fundamental frequency estimation

In inverse filtering in the linear prediction method, most of the vocal tract characteristics are filtered out into the predictor coefficients. The residual signal, the output of the inverse filter, still contains the information on the excitation source. A typical residual signal is shown in Figure 8. It is seen that large errors synchronous with pitch periods occur. A typical approach to computing the periodicity from this kind of waveform is to compute the autocorrelation function as shown in Figure 9. Two conspicuous spikes are found in the autocorrelation function, one at the origin and one at a distance of one pitch period from the origin. The fundamental frequency is then given by the reciprocal of the pitch period.

#### Problems in fundamental frequency estimation

It has been shown that the linear prediction method is quite efficient and effective for estimating the formant frequencies. However, how accurate and reliable the extraction of fundamental frequency is is an intriguing question, since there are many other techniques for estimating the fundamental frequency. Rabiner et al. (1976), in their study of the comparative performance of several pitch detection algorithms, point out the following major problems in detecting the fundamental frequency: (1) glottal excitation is not perfectly periodic; (2) defining the exact beginning and end of each period is difficult; (3) the distinction between unvoiced portions and low level voiced portions is difficult; (4) there is an interaction between the vocal tract and the glottal excitation.

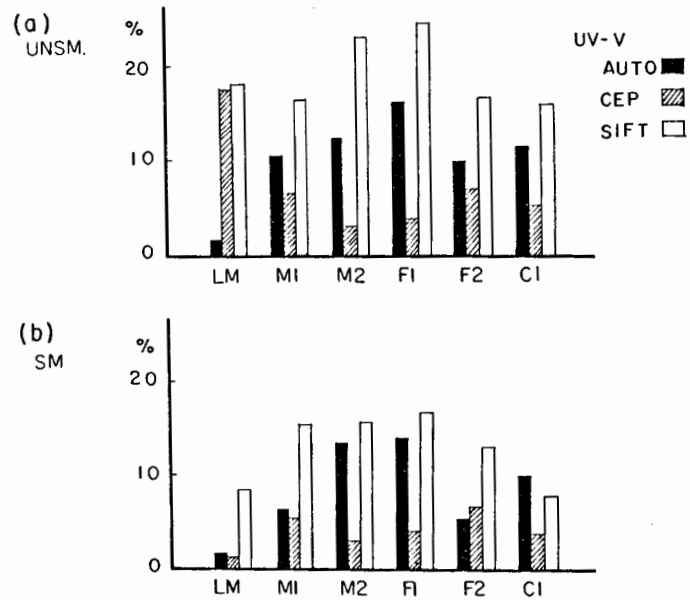


Figure 11. Unvoiced-to-voiced error for three pitch detection methods: (a) unsmoothed; (b) smoothed. (LM: low-pitched male; M1, M2: males; F1, F2: females; CI: child). The ordinate shows the percentage error rate against total number of unvoiced intervals.

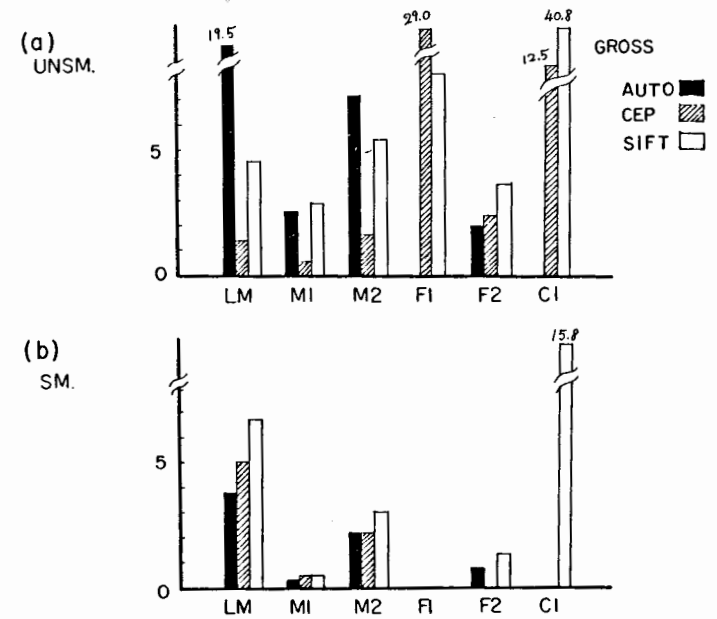


Figure 12. Gross errors for three pitch detection methods: (a) unsmoothed; (b) smoothed. (LM: low-pitched male; M1, M2: males; F1, F2: females; CI: child). The ordinate shows the average number of samples.

The above problems are intrinsic in any of the pitch detection methods. However, evaluation of several pitch detection methods indicates some differences in their performance.

#### Accuracy in fundamental frequency estimation

Let us take the following pitch detection methods from the study by Rabiner et al. (1975): (1) autocorrelation method with clipping (time domain method); (2) cepstrum method (frequency domain method); and (3) linear prediction 'SIFT'<sup>1</sup> method (time-frequency method). The types of errors can be categorized into (a) voiced-to-unvoiced error, (b) unvoiced-to-voiced error, (c) gross error in which the error in detecting the pitch period is greater than a certain threshold; and (d) fine error in which the error in detected pitch period is less than the threshold.

The above three methods were tested against six speakers (3 males, 2 females, and a child) by using four monosyllabic non-sense words and four sentences. The analysis results were compared with the standard pitch contours which were carefully measured by using a semi-automatic pitch detector. The results for the first three types of errors are shown in Figures 10, 11, and 12. The results are shown both for unsmoothed (raw data) and smoothed cases. In the smoothed case a nonlinear smoothing technique was applied to the raw data (Rabiner et al., 1975). It is seen that the nonlinear smoothing generally improves the accuracy; particularly, the gross errors are substantially improved. It is also seen that all three methods are somewhat speaker dependent. For the voiced-to-unvoiced errors, the error rate of the cepstrum method is much higher than the others except for the child speaker. For the unvoiced-to-voiced errors, on the other hand, the error rate of the cepstrum method is better than the others except for one of the female speakers for the smoothed case. In overall performance evaluation, there seems to be not much difference between the performance of the autocorrelation and linear prediction methods, except that the linear prediction method resulted in an exceedingly poor performance for the child speaker for the unvoiced-to-voiced and gross errors.

#### Other related topics

The filter box in the linear prediction model in Figure 1 contains the contribution from the glottal characteristics and the radiation effect at the lips as well as the vocal tract

-----  
1) Simplified Inverse Filter Tracking

characteristics. Since the model assumes a linear system, those factors can be separated and changed in order as shown in Figure 13. If the glottal and radiation characteristics can be eliminated by a proper preprocessing of the speech, the true vocal tract characteristics can be obtained by the linear prediction method. One of the important features of the linear prediction method is that in computing the prediction coefficients, another parameter which is called "reflection coefficient" (or "k-parameter", or "PARCOR coefficient") is obtained. A set of reflection coefficients obtained for a given speech segment gives an acoustic tube shape which has a frequency characteristic identical to the vocal tract characteristics extracted from this speech segment. In this case, the acoustic tube is represented by a concatenation of cylindrical sections of different cross-sectional areas. A reflection coefficient is defined at the boundary between two neighboring sections. Consequently, if the analysis conditions are properly chosen after preprocessing sampled speech to eliminate the glottal and radiation characteristics, the acoustic tube representation thus obtained is expected to be a good approximation to the vocal tract area function which denotes the cross-sectional areas along the vocal tract from the glottis to the lips (Wakita, 1973, 1979).

Another interesting topic is the use of the linear prediction parameters for speech synthesis. The synthesizer could be the synthesis part of the linear prediction analysis-synthesis telephony (see Markel and Gray 1976; Wakita 1976). Since the formant frequencies and bandwidths constitute the roots of the inverse filter transfer function, they can be related to the filter coefficients. The reflection coefficients, which give an acoustic tube representation of the vocal tract, are also related to the filter coefficients in the mathematical formulation of linear prediction. Thus, those parameters mentioned above are interchangeable for each other, and any of these parameters can be used for the linear prediction synthesizer.

#### Application examples

The linear prediction method has mainly been used in the area of analysis-synthesis telephony. The method is particularly effective for low bit-rate speech coding. However, the technique is equally useful for acoustical analysis of speech. In concluding this tutorial paper, several examples taken from the author's past studies will be given below.



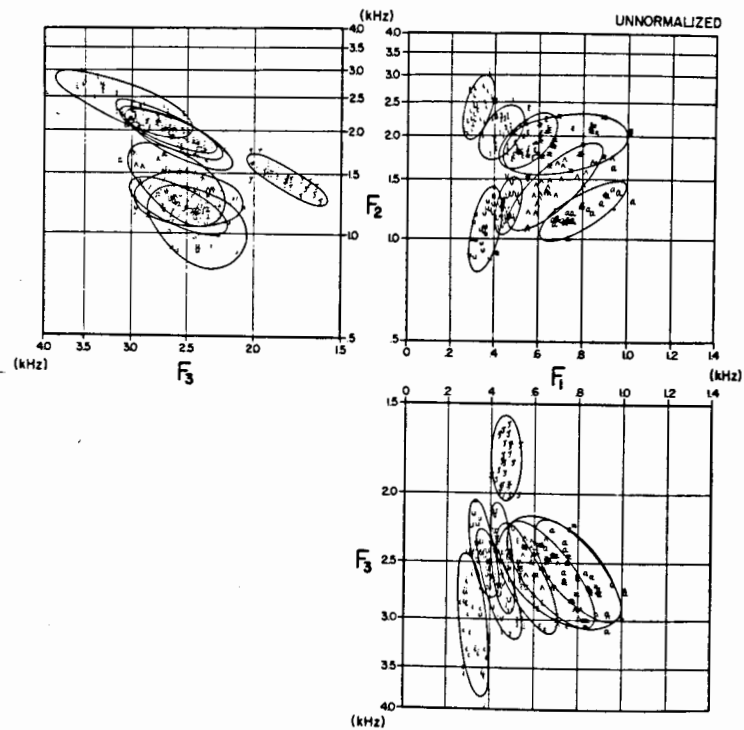


Figure 15. Distribution of formant frequencies projected onto the  $F_1$ - $F_2$ ,  $F_1$ - $F_3$ , and  $F_2$ - $F_3$  planes for 26 speakers (14 males and 12 females). Ellipses represent two standard deviations.

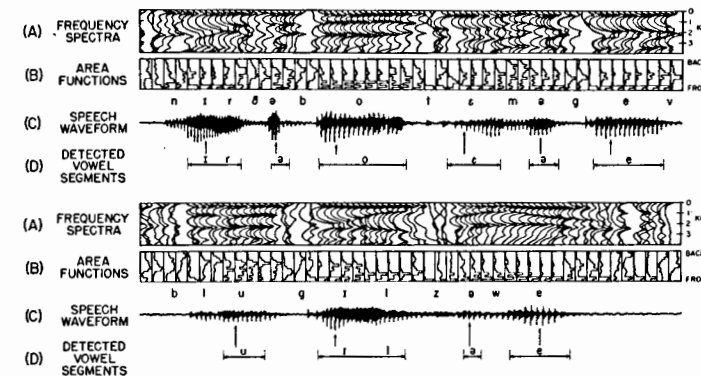


Figure 16. An example of segmenting the vowel-like intervals for the sentence "Near the boat, Emma gave blue-gills away."

### Conclusion

The concept and evaluation of the linear prediction method were described in this paper. Because of its tutorial nature, the descriptions in some cases may be inadequate from the theoretical point of view. Readers interested in more advanced knowledge are encouraged to read the original papers or other materials listed in the references.

### Acknowledgement

The author would like to thank Dr. P.-A. Benguerel, The Phonetics Laboratory, University of British Columbia, Canada, for his collaboration in investigating the use of an artificial larynx buzzer.

### References

- Atal, B. and M.R. Schroeder (1967): "Predictive coding of speech, Proc. 1967 Conf. Commun. and Process., 360-361.
- Broad, D.J. and H. Wakita (1978): "A phonetic approach to automatic vowel recognition", in *Bolc Speech communication with computers*, 52-92, London: Macmillan.
- Chandra, S. and W. Lin (1974): "Experimental comparison between stationary and nonstationary formulation of linear prediction applied to voiced speech analysis", *IEEE Trans. ASSP-22*, 403-415.

- Itakura, F. and S. Saito (1966): "A statistical method for estimating speech spectrum", Technical Report 3107, Electrical Commun. Res. Lab., NTT.
- Kasuya, H. and H. Wakita (1979): "An approach to segmenting speech into vowel- and nonvowel-like intervals", IEEE Trans. ASSP-27, 319-327.
- Makhoul, J. (1975): "Linear prediction: a tutorial review", Proc. of IEEE vol. 63, 561-580.
- Markel, J.D. and A.H. Gray (1976): Linear prediction of speech, New York: Springer.
- McCandless, S.S. (1974): "An algorithm for automatic formant extraction using linear prediction spectra", IEEE Trans. ASSP-22, 135-141.
- Rabiner, L.R., M.R. Sambur and C.E. Schmidt (1975): "Applications of a nonlinear smoothing algorithm to speech processing", IEEE Trans. ASSP-23, 552-557.
- Rabiner, L.R., M.J. Cheng, A.E. Rosenberg and C.A. McGonegal (1976): "A comparative performance study of several pitch detection algorithms", IEEE Trans. ASSP-24, 399-418.
- Wakita, H. (1973): "Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms", IEEE Trans. AU-21, 417-427.
- Wakita, H. (1976): "Instrumentation for the study of speech acoustics", in Lass (ed.) Contemporary issues in experimental phonetics, 3-40, New York: Academic Press.
- Wakita, H. (1977): "Normalization of vowels by vocal-tract length and its application to vowel identification", IEEE Trans. ASSP-25, 183-192.
- Wakita, H. (1979): "Estimation of vocal-tract shapes from acoustical analysis of the speech wave: the state of the art", IEEE Trans. ASSP-27, 281-285.

## DISCUSSION

Gunnar Fant, Wiktor Jassem and René Carré opened the discussion.

Gunnar Fant: I think that at the moment LPC analysis is more useful for communication engineering purposes, but it is certainly gaining importance in phonetic analysis: the fact that you can re-synthesize speech with rather good quality with LPC methods is a great advantage in synthesis, and LPC also makes it possible to manipulate e.g. fundamental frequency, independently of other parameters, which makes it well suited for prosodic investigations.

Formant frequencies and bandwidths describe the vocal filter, but what about the vocal source? In LPC analysis, it is treated as a constant function, more or less, but in the future we should pay more attention to the time dynamics of the source, to obtain valuable information for prosody studies. We should make dynamical matches not just to formants but also to source characteristics. (This we can do at present by carefully scrutinizing period after period of the signal, extracting presumed vocal source characteristics.) The fact that LPC is confined to an on/off, or voiced/voiceless, distinction creates some undesirable compensation effects: to compensate for a more steeply falling voice source spectrum, like we get e.g. in open syllables, the system will increase the bandwidths somewhat, which can give a consonantal effect.

Another critical problem is assessing formant frequencies with high pitched voices and in cases where  $F_0$  and  $F_1$  are close together, which is problematic in any kind of analysis.

Hisashi Wakita: mentioned a comprehensive LPC analysis of 900 vowels by a female speaker (30 vowels x 30 repetitions) where (50) unlikely analysis items were discarded by visual inspection of the vowels in  $F_1$ - $F_2$ , and  $F_1$ - $F_3$  plots [see "Application Examples", Example 1 in Hisashi Wakita's paper], but admitted that we do not yet have valid data that tell us how accurately we can estimate formant frequencies, especially when  $F_0$  and  $F_1$ , or two formants, are close together.

If we analyse a little more than one pitch period, using a very small time window and the covariance method we can, from the error signal, determine that point where the interaction between sub- and supraglottal systems is minimum (corresponding to the

closed glottis portion), and if the signal has been carefully recorded, directly from the microphone into the computer storage, so as to avoid phase distortion, we can fairly well recover the glottal wave shape from this portion.

Wiktor Jassem: What is the perspective for phonetics of these methods? First, there is the segmentation problem which can probably be solved, as suggested by professor Fant and others, by determining the maximum rate of change of the spectrum and of the time function. Secondly, there is the extraction of parameters: those extracted for automatic analysis need not be identical to those used by a human being. Thirdly, there is the problem of normalizing for individual speaker characteristics. The fourth problem is concerned with the identification of entities, which is an intricate one, because we do not know how many entities there are. The theory is that they should be sufficient to specify the output in such a way that synthesizing it we would get a normal native accent. The perceptual experiments needed to settle the question are not simple, because the adults' responses will be heavily influenced by phonemic considerations, and with very young children there will be great psychological problems. Fortunately, mathematical methods are developing that will allow us to determine, given a number of data, how many objects or entities we are dealing with. What I want to point out is that if we can get the computers to do phonetic transcriptions they will be better than transcriptions by a human being because they will be more objective.

René Carré: There are two kinds of work in speech analysis. One is the analysis of a small number of speech sounds. Formant frequencies are no problem, but to determine bandwidths we need to consider pre-emphasis, the order of the predictors, the analysis window, and the magnitude of the prediction error. All these operations take time, and such a procedure cannot be adopted in the other kind of study, of a large corpus, where a (semi-)automatic procedure has to be set up. It seems that in that case the procedure must be normalized. Is the autocorrelation method accurate enough for bandwidth measurements? Must we change (automatically or not) the order of the predictor to adapt the system to the speech sound under analysis, e.g. to nasalized vowels? What sampling rate shall we choose? How many frames should be analyzed? And so on. Finally, among the set of pole values we have to choose (automatically or not) the right formants.

Hisashi Wakita: The RMS-function is generally not sufficient to segment a chain into vowel-like and non-vowel-like sounds. But from the pseudo vocal tract area function, generated by the LPC analysis, we can calculate the ratio of the volume of the back (pharyngeal) cavity to the total volume of the vocal tract and this will generally tell us whether a segment is vowel-like or not. It will detect nasal consonants which is difficult to do from the waveform: LPC does not assume any nasal tract, but does produce a sort of equivalent acoustic tube representation, and nasal segments are fairly well detected from the back-to-total ratio of that tube.

We have also worked on the elimination of inter-speaker variability, which is of interest not just to automatic speech recognition, but also in acoustic phonetic studies of e.g. the vowel systems of languages. With LPC we can estimate the vocal tract length for each speaker and each vowel category (tract length is not constant over different vowel qualities), and then normalize to a certain length, e.g. 17 cm, a normalization which reduces the overlap in F1-F2, and F1-F3 plots and results in compact vowel distributions.

Adrian Fourcin: The LPC system represents the complexities of the vocal tract and its excitation by an exceedingly simple model: a vocal tract with no side-branches and a sharp impulse for an excitation, and yet it produces speech of very high quality. When we synthesize we have to pay attention to the zeros introduced by nasality, and the time dependence of the excitation function is also apparent if we have a standard model of the vocal tract. Is there something that we can learn from this with regard to how we hear speech?

If we knew when the point of excitation occurred and for how long a time the glottis is closed, to what extent would you be able then to improve the phonetic utility of the LPC analysis?

Hisashi Wakita: The ear is insensitive to spectral zeros, and a model which has poles and zeros in it (which is much more complicated computationally) does not perceptibly improve the quality of the speech. I have run an experiment, where various musical instruments as well as speech were passed through an artificially generated pole-zero system, and it turned out that the ear was insensitive to dips in the spectrum as large as 35 dB

(a fact which explains why HiFi loudspeakers may have even very sharp dips).

If we can determine that segment of speech where the glottis is closed, i.e. the force-free oscillations, we can apply the covariance method, which assumes that the speech waves can be approximated by a combination of damped sinusoids, and thus compute the exact vocal tract characteristics.

Gunnar Fant: A reply to Dr. Fourcin is that LPC speech sounds good because it resembles natural speech, although its source and transfer functions do not resemble those of real speech. The source function is stylized, but then there is a compensation in terms of the transfer function chosen to get the overall result correct (something which invalidates the data we get on formant frequencies and bandwidths).

Another characteristic of LPC analysis is that all the losses are concentrated at the glottal end of the system. How much does that invalidate the bandwidth data?

Hisashi Wakita: It is true that the LPC method approximates the spectral envelope, without any regard to formant frequencies and bandwidths. All the energy losses are lumped into one single resistance at the glottis end. By means of this single resistance we represent all the bandwidths of the spectrum. If we want to relate it to a particular speech production model, in terms of formant frequencies and bandwidths, it is quite useless, I think, so either we have to build more realistic models, both production and inverse transform models, or we can try to relate the simple LPC model to a more realistic, complicated model.

John Clark: There seems to be no great difference in the intelligibility levels quoted in the recent literature for predictor coded and formant coded speech. For formant coded speech, some of its phonetic weakness appears (when tested with CV-nonsense syllables) in the fricatives. Is this also the case for predictor coded speech, and what sort of evaluation have you done of the perceptual weaknesses of the system as a means of synthesizing speech?

Hisashi Wakita: Normally, with the LPC analysis-synthesis we use the extracted coefficients as they are, but we replace the residual signal with a pulse train which makes the voiced/unvoiced decision very critical, and missing just one frame can be per-

ceptible. We can, however, restore the original signal by using the residual signal for excitation. For phonetic evaluation purposes I think we have to choose the excitation source carefully, - maybe not the residual signal itself, but one with which we do not lose too much information about the source.



SYMPOSIUM NO. 1: PHONETIC UNIVERSALS IN PHONOLOGICAL SYSTEMS AND THEIR EXPLANATION

(see vol. II, p. 5-59)

Moderator: John J. Ohala

Panelists: Thomas V. Gamkrelidze, André-Georges Haudricourt,  
Robert K. Herbert, Jean-Marie Hombert, Björn Lindblom,  
Kenneth N. Stevens, and Kenneth L. Pike

Chairperson: Bertil Malmberg

JOHN J. OHALA'S INTRODUCTION

*Phonetic universals* is such a large subject that the members of this symposium despaired of being able, in the short time allotted, to give adequate consideration to any of the general aspects of the theory or practice of the field or to solve any of its "great problems". It was decided, therefore, that the moderator would make a few brief general comments about some of these larger issues, more or less "for the record", but that most of the time of the symposium be devoted to the discussion of one very specific problem in the area of phonetic universals.

General Problems and Issues in Phonetic and Phonological Universals

(In this report I will use the shorter phrase 'phonological universals' for the longer, somewhat unwieldy expression 'phonetic universals in phonological systems', the official topic for this symposium.)

1. Before beginning this discussion, we should define what we mean by *phonological universals*. As this term has come to be used, it means *systematic patternings of speech sounds cross-linguistically*. This definition does not require that the pattern be manifested in every human language, merely that it have sufficient incidence in the languages of the world such that its occurrence could not be attributed to chance. It is assumed, though, that all languages, indeed, all human speakers, are potentially subject to whatever "forces" create these patterns, but an overt manifestation of these forces may or may not occur and if it does occur, may take different forms. For example, to consider a case discussed extensively by Professor Gamkrelidze, it is presumably the same universal factors which are responsible for the asymmetrical gap in the voiced velar stop position (/g/)

in the segment inventories of Dutch, Czech, and Thai, as are responsible for the disproportionately low incidence of /g/ in the lexicon or in running speech of many languages. Likewise, whatever causes the asymmetrical absence of /p/ in Arabic, Nkom, and Chuave, is also responsible for the limited distribution of /p/ in Japanese, i.e., it only appears intervocalically and as a geminate.

2. The concern with phonological universals in our field has both theoretical and practical consequences. Some 100 years ago our intellectual forefathers, Ellis, Sweet, Passy, Lepsius, Jespersen, and others, provided us, in the phonetic alphabet and the descriptive anatomical and physiological terms accompanying it, the equivalent of the Linnean system of classification in biology or Mendeleev's periodic table of the elements in chemistry. Today, I believe it safe to say that we have reached the stage equivalent to that which Bohr's model of the atom represented in physics and chemistry. We have a framework within which to observe, to describe, and to establish natural classes of phonetic and phonological entities and processes in all human languages. We are also able, with obvious limitations, to predict and explain the behavior of speech sounds. Commendably, in many cases, these explanations are based on empirically-supported models of parts of the speech communication process. Although it is obviously the case that as we deepen our understanding of some of the basic physical, physiological, and psychological mechanisms serving speech, we also are better able to explain many phonological universals; it is also true that in many cases *it is our observation of phonological universals which leads to a greater understanding of speech mechanisms*. The literature in phonological universals is even now causing us to critically re-examine some of the most fundamental concepts in phonetic and phonological theory, for example, the notions of 'segment', of 'distinctiveness', etc., and to explore in considerable detail in the laboratory basic acoustic, aerodynamic, and auditory mechanisms in speech.

In the practical realm phonological universals can aid us in the analysis and understanding of the phonologies of individual languages: they tell us what to look for and they help us to choose alternative scenarios for the history of sound changes in the language. I personally believe that phonological universals

can also aid us in such cases of *applied phonology* as speech synthesis, automatic speech recognition, speech pathology, speech therapy, and language teaching. It must be said, however, that at present there has been very little penetration of universals in these areas.

3. Phonological universals are found in many different forms, e.g., segment inventories, segmental sequential constraints ("phonotactics"), allophonic variation, sound change, morphophonemic variation, dialect variation, patterns of sound substitution by first and second language learners, frequency of occurrence of sounds in the lexicon and in connected speech, conventional and esthetic use of speech sounds in onomatopoeia, poetry, jokes, singing, etc. Can we bring all of these disparate phenomena under one theoretical umbrella, using one of these as the base or primitive from which the others may be derived, or, possibly, deriving them from some separate principle external to all of them?

4. Another general issue concerns the problem of how to obtain a truly representative sample of sound patterns from a variety of languages such that the sample is not biased by including too many or too few languages having certain genetic, typological, or geographical linkages. The many pitfalls of attempting a quantification of phonological data from large samples has been discussed previously, including such concerns as how one differentiates a language from a dialect, whether one should look at the behavior of phones or phonemes and if phonemes, whose conception of the phoneme, etc? The fact is, most works on phonological universals ignore this issue and seem to rely on the investigator's intuitive "feel" for what constitutes a proper sample. Is there any way to make this process objective? How can we create an unbiased sample; how large should it be?; what criteria should we apply in admitting a language to the sample? Once we have the supposedly unbiased sample, what type of statistical analysis should we apply to it in our attempts to prove or disprove universal tendencies?

My own solution to this problem, a solution which has parallels in other scientific disciplines, is to make sure that any posited universal is supported both *inductively* -- that is with lots of examples (and few counterexamples) -- and *deductively* --

that is, by what we know to be the underlying operating principles of speech production and perception.

5. A related issue is whether or not some of the claims made about phonological universals may be distorted by observer bias, i.e., be self-fulfilling prophecies. It has been claimed, for example, that all languages code speech in terms of phonemes. But I know of no universally-accepted algorithm which discovers phonemes. And if there were, do we now have any evidence that phonemes and all the properties attributed to them, have psychological and/or physical reality?

A very clear example of the perils of observer bias surrounds claims about universals of syllable structures. It has been claimed that within a syllable, one should not find a transition from voiced to voiceless to voiced. Upon being presented with an apparent counterexample such as [itv], the claimant would protest that there is a syllable boundary between the [t] and [v]! The potential for similar circularity enters into any claim which contains terms that cannot be objectively defined. And this, unfortunately, is true of a very large number of terms used in phonetics and phonology, including terms such as consonant, vowel, segment, syllable, sonority, strength, lenition, etc.

Would we find a different set of universals if we adopted the parallel, hierarchic system such as Professor Pike advocates? Would we have a different, more interesting set of universals if we included in the description of sounds, as Professor Stevens proposes, the sensory information each sound gives rise to?

#### A Specific Problem in Phonological Universals

The problem selected for special attention during this symposium is by no means a small one and it is doubtful that it will be solved very quickly, certainly not in the short time allotted us. Nevertheless, it is a problem that intersects with the particular interests of most members of the symposium and is a matter to which many members of the audience can contribute. The problem is stated in a deliberately provocative way in order to stimulate discussion.

The notion of a vowel "space" has been used in phonetics for about 2 centuries but it is only recent evidence which points to this space having acoustic-auditory correlates. The research of Lindblom and his colleagues suggests that the placement of vowels

in this space in various languages is dictated by the principle of maximal perceptual difference, i.e., that however many vowels there are in the system, they tend to arrange themselves in the available space in such a way as to maximize their distance from each other. This principle seems to adequately predict the arrangement of systems with approximately 7 or 8 vowels. It would be most satisfying if we could apply the same principles to predict the arrangement of consonants, i.e., posit an acoustic-auditory space and show how the consonants position themselves so as to maximize the inter-consonantal distance. Were we to attempt this, we should undoubtedly reach the patently false prediction that a 7 consonant system should include something like the following set:

d, k', ts, ʔ, m, r, ʒ.

Languages which do have few consonants, such as the Polynesian languages, do not have such an exotic consonant inventory. In fact, the languages which do possess the above set (or close to it), such as Zulu, also have a great many other consonants of each type, i.e., ejectives, clicks, affricates, etc. Rather than maximum differentiation of the entities in the consonant space, we seem to find something approximating the principle which would be characterized as "maximum utilization of the available distinctive features". This has the result that many of the consonants are, in fact, perceptually quite close -- differing by a minimum, not a maximum number of distinctive features.

Does this mean that consonant inventories are structured according to different principles from those which apply to vowel inventories? Could it mean that the "spaces" both consonants and vowels range in, are limited by the auditory features (= parameters) recognized by the particular language? Or does it mean that we are asking our questions about segment inventories in the wrong way?

#### COMMENTS FROM THE PANELISTS

K.N. Stevens: In an acoustic representation of connected speech we find certain regions where there are rapid (10-30 msec) changes in a number of acoustic parameters, e.g., amplitude, periodicity, and spectrum. A hypothesis that has emerged from our and Chistovich's research, is that the attention of the listener is drawn

to these regions, more so than to other regions where changes are less rapid. These regions are, first of all, markers of consonants, but additional information can also be packaged in them along several orthogonal dimensions. We believe languages therefore tend to "select" a consonant inventory that uses up most of these dimensions. These primary dimensions are: [+ voice] (presence/absence of periodicity), [+ nasal] (presence/absence of low-frequency murmur), [+ continuant] (unbroken/interrupted sound), [+ grave] (low-/high-frequency tilt to the spectrum), [+ compact] (energy spread out/concentrated). After processing the information in these regions of rapid change (= high rate of information transfer), the listener's attention may focus on the remaining regions and here lie the cues for such dimensions as palatalization, pharyngealization, clicks, etc. It logically follows that the learning of (or introduction of) such distinctions will follow the learning of distinctions coded in the regions to which primary attention is directed.

B. Lindblom: We have recently followed up and improved on our early work on predicting vowel inventories and I think the research strategy we have used could be applied to consonant inventories, too. Briefly, our procedure is to 1) specify a physiological model of the vocal tract and use it to define 2) the range of humanly possible vowels and from this derive 3) the (universal) human acoustic vowel space, a continuum, and, finally, 4) to employ an auditory model to define a perceptual space to accommodate a specified number of vowels. The last step consists of convolving an input power spectrum (of a given vowel) with an auditory filter derived from masking data, thus yielding a hypothetical auditory excitation pattern. We assume that, other things being equal, the probability of any two vowels being confused, that is, their perceptual closeness, will be related to the overlap area enclosed by their excitation patterns. We believe vowel systems evolve so as to make vowel identification efficient and this is done by making perceptual differences between vowels (quantified as mentioned above) maximally or, perhaps, sufficiently large. This new measure of perceptual distance yields much more reasonable predictions about vowel placement; in particular, it eliminates the excessive number of high central vowels that plagued previous models.

A preliminary typological study of diphthongs shows that [a<sup>i</sup>]

and [a<sup>u</sup>] are the most favored. This result is compatible with the new properties of our model's perceptual space and provides evidence for a principle of perceptual differentiation applying not only paradigmatically, but also sequentially. Consonant inventories can be studied within a paradigm such as this.

K. Pike: My own approach to phonetic analysis is a bit different from that of most of my fellow panelists. Although I have often been helped by acousticians when I have brought my phonetic problems to them, I would rather argue that the reductionism, so necessary in the laboratory, is detrimental to linguistic analysis in the field. I can illustrate this with an examination of a short poem by E.E. Cummings. [Text and detailed commentary omitted.] Although one can point out puns, details of orthography, prosody, and even cultural allusions which contribute to the overall effect, the poem, like language, functions as a whole. I am encouraged by the enlarged scope of phonological inquiry demonstrated at this congress, e.g., the work on syllables. The study of vowel spaces should also be enlarged to include what I call 'pharynx space' (changes in vowel quality by modifications of pharyngeal width and larynx height) and by taking into consideration the psychological reality of vowel structure.

J.-M. Hombert: A surprising number of people I have met at this congress are quite skeptical about the existence of phonological universals. Although one can cite countless examples of cross-language similarities in sound inventories, sound changes, and phonological processes, there are, of course, always counterexamples to almost any generalization one might make. Perhaps the answer to this is to pay more attention to the diachronic aspect of universals: the counterexamples may just be unstable transitional states between more natural states. Moreover, it is often possible to find that certain cited counterexamples cease to be so if one looks into the details more closely, e.g., in cases of tonal development from obstruents, a voiced stop giving rise to a high tone runs counter to the usual patterns, but if it was found that the voiced stop had first become an implosive, an expected development, then the case is no longer a counterexample.

Concerning the sampling problem, mentioned by the moderator, it is particularly acute in the case of perceptual data. This can be solved if we start discovering ways to take our laboratories in-

to the field and thereby gather perceptual data from a wide variety of languages.

R. Herbert: A consideration of the factors constraining the introduction into a consonant inventory of complex sound types, e.g., affricates, pre- and post-aspirated consonants, and especially pre-nasalized consonants, may provide insight into the constraints on consonant inventories as a whole. Obviously, the parts of such complex segments must be sufficiently different from each other so that they may both be perceptually salient within the time span of a single segment, e.g., the nasal/oral distinction used in pre-nasalized stops. It must also be possible to articulate the parts within this same time span. Thus there are limits on the number of components in single segments: usually 2, but 3 in the case of pre-nasalized affricates, and rarely more. Most such complex sounds involve at least quasi-homorganic components, and thus nasal and stop combinations are frequently encountered but lateral and stop combinations less so since laterals, unlike nasals, have limited capacity for homorganicity. We might also speculate that the relative ordering of the components in complex segments is governed by the same factors that determine optimal syllable codas: the first element is generally the more common syllable coda, it being understood that optimal syllable codas are drawn first from the opposite ends of the sonority hierarchy, e.g., glides, nasals, [ʔ], and voiceless stops, before involving segment types from the middle, e.g., laterals, voiced stops, fricatives.

A.-G. Haudricourt: The search for phonological universals seems to me to be like the quest for the philosopher's stone. As for phonetic changes, it is more profitable to look at the conditions for the appearance of the phenomena rather than for their existence. Language is a social phenomenon and one of its main functions, communication, causes the development of new phonemes. Sindhi provides an example: its whole series of voiced stops, when long, has become preglottalized in order to remain distinctive. Language also has a socio-ethnic function and so preglottalization may appear without any phonological conditioning, as happens in Vietnamese and the Henan dialect of Chinese. In these cases, one or two preglottalized consonants are sufficient for the social function and it is normal that they should be the easiest to articulate (β, ɖ). Likewise, preglottalized consonants can disappear for a variety of rea-

sons. The loss of these sounds in Vietnamese was in part due to the presence of tones (which made the voicing superfluous) but has also been aided by the sociolinguistic environment in, e.g., Saigon. These facts are outside the domain of instrumental phonetics.

T.V. Gamkrelidze: I believe an understanding of the principles governing the structure of consonant and vowel inventories will come from typological phonology and experimental phonetics. An important task for typological phonology today is the establishment of constraints or relations of markedness or dominance between certain bundles of co-occurring features. For example, as detailed in the printed version of my paper, in the subsystem of stops and fricatives, [+voice +labial] is dominant (unmarked) with respect to the co-occurring features [+voice +velar]. Thus, among voiced stops, /b/ is dominant, /g/ is recessive. Also, among voiceless stops, /k/ is dominant, /p/ is recessive. These relations stem from the specific acoustic and articulatory properties of the features involved. In the examples mentioned, the volume of the air chambers plays a part. Gaps in the paradigmatic system of obstruents will generally reflect these dominance/recessiveness relations. These relations can therefore help us to better understand sound change and to do language reconstruction more realistically. In light of this, the classical reconstruction of the Indo-European occlusive phonemes appears to be linguistically improbable in that (among other things) it assumes the series with the missing labial were voiced stops. Reinterpreting this series as ejectives brings the IE obstruent system into full conformity with typological studies.

J.J. Ohala: I would speculate that a universal vowel and consonant space does not exist. Each language "chooses" some restricted set of features or dimensions for these spaces. It is common knowledge, for example, that a native speaker of one language is 'deaf' to certain features used in other languages. It is true that the Lindblom model does have a remarkable degree of success in predicting the structure of systems with a small number of vowels. But it is significant that it breaks down when a large number of vowels are involved, very likely because one or two dimensions other than those used in the model are also involved, e.g., vowel duration, diphthongization, voice quality. It could be that vowel spaces, unlike consonant spaces, have rather few possible dimensions and that most languages make some use of the most salient dimensions (those based

on spectral shape). In consonant systems, it is well known that there are more possible dimensions to choose from and so the discrepancy between reality and the predictions of a maximum-perceptual-distance model are more evident. Thus, the differences between vowel and consonant systems in this respect are only apparent. What is more remarkable -- to me, at least -- is the highly symmetric nature of consonant proliferation. The mechanism of proliferation is reasonably clear, e.g., stop plus [ʔ] yields a glottalized series of stops or ejectives, but why should proliferation almost always yield a whole new row or column of such consonants?

## DISCUSSION

K.N. Stevens: It is true, as Professor Gamkrelidze notes, that aerodynamic factors contribute to the asymmetries in obstruent systems, but auditory factors are important, too. The noise or burst of a voiceless velar will give a very clear indication of compactness -- more so than a voiced velar, whereas a voiced labial will reveal the feature [+grave] better than a voiceless labial. J. Ohala and K. Stevens discussed the need, in the search for the most salient auditory dimensions, of finding the perceptual cues for such striking sounds as ejectives.

K. Pike and J. Ohala mentioned specific instances of vowel and consonant systems utilizing voice quality as a distinctive dimension, e.g., certain languages of Nepal, various Nilotic languages, Korean, Javanese, Cambodian, Gujarati.

B. Lindblom: It is possible, in principle, to include other dimensions in the vowel space, but it is better at this stage of research to make our models precise and quantitative. At present then, it is better to restrict the investigation to spectrally-based dimensions. I agree with Ohala that listeners react to vowel stimuli in language-specific ways. In fact, some of our own research shows that Swedish listeners put more subjective distance between the vowels in the crowded front region of the Swedish vowel space than would have been predicted by our model's spectrum-based metric. But let us not be too hasty in discarding the notion of a universal vowel space. After all, this may be what the child brings to the language-learning task.

J. Ohala: I concede that I overstated my position. There undoubtedly is a universal vowel space and each language chooses a sub-

space within it. No doubt there is some order according to which features are chosen first.

T.V. Gamkrelidze: The greater proliferation of consonants as opposed to vowels is due to the greater number of possible dimensions in consonant systems. In theory, of course, an infinite number of vowels could be produced, but practically the number is small due to auditory and articulatory constraints.

A. Haudricourt: (In response to a question from J.-M. Hombert) The search for phonological invariants and for culture-specific phenomena is not incompatible, but they are two different problems. First we must investigate the *function* of language and only then look at its phonetic realization.

B. Lindblom: Given the well known *discreteness* of language, it might be asked why, in our model, we start with a *continuous* vowel space. The answer is that we do not yet have a theory that predicts that language should have discrete units such as distinctive features. The theory of distinctive features we do have is based on induction. I think the discreteness has to be deduced or derived as a consequence of more fundamental principles. Even so, a totally discrete model will still not explain why, in languages with few vowel contrasts, the extreme corner vowels tend to be phonetically less extreme (as noted by Crothers).

(To Prof. Stevens:) The quantal phenomena you find in the articulatory-to-acoustic transformation cannot be the only source of phonological discreteness. Surely, memory mechanisms must be involved as well (cf. the work of G. Miller and I. Pollack on elementary auditory displays).

K. N. Stevens: I agree with all of your points. I would just say that in the vowel space there are some regions which are more stable (or discrete) than others in that a wide range of articulations would give rise to the same acoustic signal. So the vowels will be within these regions, the exact location determined by factors such as your model incorporates. It is possible, too, that the whole space may shift in one direction or another due to different so-called 'basis of articulation' of various languages.

B. Lindblom: Isn't this a denial of the possibility for a universal framework?

K.N. Stevens: I don't think so. I view these shifts as being fairly small. The high front vowels in various languages may not be phonetically identical, but they are still high front vowels.



K. Pike: It won't work to say it is either 'discrete' or 'continuous'. We need 'particle' or 'wave' descriptions, both of which are observer-related, and a 'field' view which describes it in terms of an overall system.

C.J. Bailey and T.V. Gamkrelidze expressed differing views on how much weight to give to typological evidence as opposed to comparative (within-family) evidence when doing reconstructions.

C. Scully: A propos of pre-nasalized stops, I have found in air-flow traces that the velum closes very late during the closure portion of post-pausal voiced stops, almost as if some aspects of speech are begun while certain acts of respiration (open velum) are still in play. This may be a good example of a mechanically determined feature of pronunciation that might become generalized and taken up as a linguistic feature.

S. Anderson: I wish to take issue with the assumptions (or by Chala, an explicit proposal) that claims about phonological structures must be verifiable in terms of substance in some other domain, typically phonetic. At the Phonology session of this congress I sketched a rather different approach to phonology which assumes that there is a systematic domain which is relevant to the nature of language but which isn't directly reducible to other domains. According to this view, the facts that are directly susceptible of phonetic explanations are, in a sense, exactly what is irrelevant to phonology.

F. Longchamp: (To Hombert) You haven't made a clear case for the decreased saliency of the centralized vowels. The vowels that behaved oddly in your study seem to be the one-formant vowels. Of course, subjects can give labels to these vowels but this may have no relevance to natural speech.

H.-H. Jeng: I think child language studies can provide evidence relevant to the questions on the elaboration of segment inventories. In the early speech of my son the consonant system used only the features for t stop and those for different places of articulation. Later on, features were added to differentiate nasality, aspiration, frication, etc. In the case of vowels, only height features were used at first. Later, front-back and rounding were differentiated. I think these early segment systems represent the universal core upon which further elaborations of the system can be built.

N. Waterson: I question the phonemic basis used in work on universals. There is much evidence that the proper domain of many phonological processes is something more like the word. In sound change the position of the sound in the word and its phonetic context is very important. Children will often produce the correct degree of vowel openness in vowels in a 2-syllable word but not the correct frontness or rounding feature. Thus, when looking for universals we should look for patterns in the domain of the whole syllable or word.

H. Andersen: I don't see how Lindblom's model will accommodate vowel mergers which are very common diachronically. Nor can this problem be solved as recommended by Hombert by assigning the merged vowels to an unnatural transitional state which will eventually revert to a stable natural state. How is one to identify transition as opposed to stable state? The solution, I think, is to recognize that the vowel (as well as the consonant) space is used for more than just diacritic purposes: they also carry information about their consonant environment, about the style of speech used by the speaker as well as his age and social class membership. Thus when the vowels slide around it must be because these subsidiary functions lose their value and are re-interpreted as basic values of the vowel phonemes themselves. This notion is fully in accord with the views expressed here by Profs. Pike and Haudricourt.

L. Jacobson: I can provide some more details on the vowel systems of certain Nilotic languages (alluded to by Ohala) and and at the same time show that they are compatible with Lindblom's model. My own acoustic analysis of the 9 vowel system of Luo shows that many of the non-low vowels show great overlap in an F1 x F2 x F3 space. They can be separated, however, by adding a dimension of voice quality (or pharynx size): breathy voice vs. normal or creaky voice. When this is done, all the vowels are still maximally distant from the other vowels *on the same plane*.

I. Maddieson: It was mentioned (by Lindblom) that high vowels in systems with few vowels tend to be less peripheral. This is a crucial fact and suggests that *maximal* dispersion of entities in an auditory space isn't required. I find supporting evidence for this view in the structure of tonal spaces: words borrowed from a 2 level-tone language into a 3 level-tone language reveal that the high tone of the 2-tone language is equal to the mid-tone of the 3-tone

K. Pike: It won't work to say it is either 'discrete' or 'continuous'. We need 'particle' or 'wave' descriptions, both of which are observer-related, and a 'field' view which describes it in terms of an overall system.

C.J. Bailey and T.V. Gamkrelidze expressed differing views on how much weight to give to typological evidence as opposed to comparative (within-family) evidence when doing reconstructions.

C. Scully: A propos of pre-nasalized stops, I have found in air-flow traces that the velum closes very late during the closure portion of post-pausal voiced stops, almost as if some aspects of speech are begun while certain acts of respiration (open velum) are still in play. This may be a good example of a mechanically determined feature of pronunciation that might become generalized and taken up as a linguistic feature.

S. Anderson: I wish to take issue with the assumptions (or by Chala, an explicit proposal) that claims about phonological structures must be verifiable in terms of substance in some other domain, typically phonetic. At the Phonology session of this congress I sketched a rather different approach to phonology which assumes that there is a systematic domain which is relevant to the nature of language but which isn't directly reducible to other domains. According to this view, the facts that are directly susceptible of phonetic explanations are, in a sense, exactly what is irrelevant to phonology.

F. Longchamp: (To Hombert) You haven't made a clear case for the decreased saliency of the centralized vowels. The vowels that behaved oddly in your study seem to be the one-formant vowels. Of course, subjects can give labels to these vowels but this may have no relevance to natural speech.

H.-H. Jeng: I think child language studies can provide evidence relevant to the questions on the elaboration of segment inventories. In the early speech of my son the consonant system used only the features for + stop and those for different places of articulation. Later on, features were added to differentiate nasality, aspiration, frication, etc. In the case of vowels, only height features were used at first. Later, front-back and rounding were differentiated. I think these early segment systems represent the universal core upon which further elaborations of the system can be built.

N. Waterson: I question the phonemic basis used in work on universals. There is much evidence that the proper domain of many phonological processes is something more like the word. In sound change the position of the sound in the word and its phonetic context is very important. Children will often produce the correct degree of vowel openness in vowels in a 2-syllable word but not the correct frontness or rounding feature. Thus, when looking for universals we should look for patterns in the domain of the whole syllable or word.

H. Andersen: I don't see how Lindblom's model will accommodate vowel mergers which are very common diachronically. Nor can this problem be solved as recommended by Hombert by assigning the merged vowels to an unnatural transitional state which will eventually revert to a stable natural state. How is one to identify transition as opposed to stable state? The solution, I think, is to recognize that the vowel (as well as the consonant) space is used for more than just diacritic purposes: they also carry information about their consonant environment, about the style of speech used by the speaker as well as his age and social class membership. Thus when the vowels slide around it must be because these subsidiary functions lose their value and are re-interpreted as basic values of the vowel phonemes themselves. This notion is fully in accord with the views expressed here by Profs. Pike and Haudricourt.

L. Jakobson: I can provide some more details on the vowel systems of certain Nilotic languages (alluded to by Ohala) and at the same time show that they are compatible with Lindblom's model. My own acoustic analysis of the 9 vowel system of Luo shows that many of the non-low vowels show great overlap in an F1 x F2 x F3 space. They can be separated, however, by adding a dimension of voice quality (or pharynx size): breathy voice vs. normal or creaky voice. When this is done, all the vowels are still maximally distant from the other vowels *on the same plane*.

I. Maddieson: It was mentioned (by Lindblom) that high vowels in systems with few vowels tend to be less peripheral. This is a crucial fact and suggests that *maximal* dispersion of entities in an auditory space isn't required. I find supporting evidence for this view in the structure of tonal spaces: words borrowed from a 2 level-tone language into a 3 level-tone language reveal that the high tone of the 2-tone language is equal to the mid-tone of the 3-tone



language, the implication being that systems with 3 tones use more of the available tone space than do those with 2 tones. We could explain all this as well as the pattern of elaboration of consonant systems by the generalization: additions to these spaces first involve pushing the boundaries of the existing dimensions and then by recruiting additional dimensions for additional contrasts.

L. Lisker: Is the search for universals a viable enterprise if we can't be sure that we are aware of all the features that human languages make use of? New ones are discovered all the time. Also, when making generalizations about segment inventories, we should be clear what we're talking about: the /g/ in English is not the same 'beast' as the /g/'s in Spanish or French, for example. The problem is that the C's and V's we count are invariably the product of the phonologist who uses other than purely phonetic criteria in deciding how to classify sounds.

H. Galton: Considering cases like Ubykh, a Caucasian language with 80 consonants and no more than 2 vowels, and English with about 1/3 as many consonants and many more vowels, I wonder if Prof. Gamkrelidze would accept the tentative universal that is there a kind of balance between a language's consonant and vowel inventories, i.e., that one develops at the expense of the other?

T.V. Gamkrelidze: The number of consonants always exceeds that of vowels since the possibilities for auditory and articulatory contrasts is greater for consonants.

J. Ohala: Regarding the relative merits of a formalist vs. a physicalist research strategy in phonology, the issue raised by Prof. Anderson, I suggest this be decided by examining the 'track record' of the two approaches in providing explanations in phonology.

Reflecting on several of the comments made here, I would suggest we consider the possibility that the single multi-dimensional perceptual space that both consonants and vowels range in is not simply defined by the various spectral features (F1, F2, F3), amplitude, periodicity, etc., but rather the first derivative --the rate of change-- of those features. R. Port at Indiana as well as Lindblom have explored this possibility. In this case, the units would no longer be phonemes as such, but rather the transitions between them. These units (more numerous than phonemes) tend to be more invariant, too.

SYMPOSIUM NO. 2: THE PSYCHOLOGICAL REALITY OF PHONOLOGICAL DESCRIPTIONS

(see vol. II, p. 63-128)

Moderator: Victoria A. Fromkin

Panelists: Lyle Campbell, Anne Cutler, Bruce L. Derwing, Wolfgang U. Dressler, Edmund Gussman, Kenneth Hale, Per Linell, and Royal Skousen

Chairperson: Bengt Sigurd

VICTORIA A. FROMKIN'S INTRODUCTION

The topic of this symposium is a controversial one. We are hopeful that the debate will lead to new insights and understanding and will help to clarify issues which are important to all sides of the argument. We expect new questions to be raised, questions which we are certain will stimulate the search for answers as to the nature of human language and speech.

Throughout this IXth Congress, the complexities of speech production and perception have been discussed. While we have learned a great deal about these phenomena in the 48 years since the first International Congress of Phonetic Sciences, we still have more questions than answers. The heart of our problem is like that of all scientists, "to explain the complicated visible by some simple invisible." (Perrin, 1914) This is the aim of theory construction, the effort to find a simple, elegant, but "true" (or as close to truth as it is possible to get) accounting of, description of, explanation for the complexities of the phenomena of interest. There is, however, no single approach to how one goes about constructing and validating a theory. That this symposium attests to such differences is revealed in the proceedings (vol. II). We do not even agree as to what constitutes a true theory. The disagreements are, of course, philosophical rather than "scientific". One side of the philosophical debate is set forth by the Nobel prize winning geneticist, François Jacob (1977):

"... the scientific process does not consist simply in observing, in collecting data, and in deducing from them a theory. One can watch an object for years and never produce any observation of scientific interest. To

produce a valuable observation one has first to have an idea of what to observe, a preconception of what is possible. Scientific advances often come from uncovering a hitherto unseen aspect of things as a result, not so much of using some new instrument, but rather of looking at objects from a different angle. This look is necessarily guided by a certain idea of what the so called reality might be."

What the reality is constitutes the subject of this symposium. In our case, the reality is a mental or psychological one. We have thus rejected as too confining an earlier definition of linguistics as a classificatory science. (Hockett, 1942) It is no longer enough for a grammar to account for the facts, i.e. the raw data, with the "maximal degree of generalization". The grammar must be a model of the internal grammar constructed by the child; only then will we provide a true description of the language, or a psychologically real grammar.

Even when there is agreement on this aim, different approaches to the job before us are taken. Some linguists and psycholinguists believe that to achieve this goal, it is necessary to test each posited rule in any descriptive grammar to see if it is truly "real". Others suggest that what we are seeking are, rather, constraints on the form of grammars, or a theory of grammar which will answer the question "what is a possible language?" This latter view suggests that with proper constraints any language specific grammar which is permitted by the theory will be psychologically real in that it would be learnable, acquirable by the child when confronted with linguistic data. We all agree that a grammar which is in principle or in fact not "learnable" cannot be psychologically real.

The psychological reality problem did not arise, nor could it have arisen, among linguists such as those who followed Bloomfield in America as they rejected any form of mentalism in linguistics. But even in the early period of the transformational/generative grammar paradigm, the period in which the notion of language as a cognitive system was reintroduced as a legitimate one, there were too few constraints placed on grammars.

I am reminded of the Schachter and Fromkin (1968) phonological analysis of Akan in which final stop consonants /p/, /t/, and /k/ are posited in lexical representation. These

voiceless stops do not surface phonetically in this context. The question that such an analysis poses is whether the Akan child language learner can hypothesize the existence of these final consonants when they never occur in any forms the child hears. Chomsky and Halle (1965) discussed this question a number of years ago.

"For the linguist or the child learning the language, the set of phonetic representations of utterances is a given empirical fact. His [sic] problem is to assign a lexical representation to each word, and to develop a set of grammatical (in part, phonological) rules which account for the given facts. The performance of this task is limited by the set of constraints on the form of grammars. Without such constraints, the task is obviously impossible; and the narrower such constraints, the more feasible the task becomes."

There are no a priori principles which can tell us what the child is capable of constructing and what she is not. We do not know what the mind is capable of, either the adult mind or the immature mind. In fact, the goal of phonological theory is to provide an answer to the questions concerning the kinds of phonological representations the child can construct, and the rules which can relate these to surface phonetic forms, if indeed there is a difference between these levels. This too is a question for which there is no a priori answer.

The task then of establishing constraints on such a theory such that it will delimit the class of possible grammars to those which are psychologically real, which can be, and which are, acquirable by at least some children, is a task facing us all. If this is the general goal for phonological theory, and let us assume it is, then the question of "psychological reality" is a non-question. We need rather to ask of a theory: "Is it correct?" not "Is it psychologically real?" Or perhaps we should say that the answer to these questions will be identical. In other words, a correct theory of grammars will be a theory of psychologically real grammars.

Unfortunately, even if we agree on this, we find disagreements as to what is meant by psychological reality. I have

therefore asked the participants in this symposium to address this question, to tell us their conception of psychologically real phonological theory.

Closely tied to this basic question are those concerned with the kinds of evidence which can be used to show the reality of a grammar, a lexical entry, an abstract segment, a rule, evidence used to validate or invalidate general theories or particular phonological analyses. In a number of the papers presented in volume II a distinction is made between "external" and "internal" evidence. "External" evidence, as I noted in my summary (p. 63-66), included acquisition data, language disturbance, borrowing, orthography, speech and spelling errors, metrics, casual speech, language games, historical change, perception and production experiments etc. (Cf. Zwicky, 1975) Internal evidence, according to those who make this separation, refers, on the other hand, to facts drawn from the grammar itself, significant generalizations, simplicity factors, distributional criteria, morphemic alternations, etc.

There are linguists, including some of the participants in this symposium, who regard external evidence as more worthy of consideration, as data to be more highly valued than internal evidence. It is not quite clear to me why this should be so. And, in fact, it has been argued that if internal and external evidence are contradictory, internal evidence should prevail. (Cf. below for discussion of Gussman's paper.) External evidence is often performance data, either elicited or observed in actual speech or perception. Speech error data are of this kind. Although I have found, in speech errors, evidence for the independence of features as shown in (1)

(1) Target: Cedars of Lebanon Error: ... Lemadon  
where only the value of the feature [nasality] is switched, Klatt (1979) finds "little evidence in the speech error corpus to support independently... movable distinctive features as psychologically real representational units for utterances." While I am not ready to concede to Klatt, let us assume, for the purpose of this argument, that he is correct. Can we conclude from this that a theory of phonology should not represent segments as bundles of features? If we did, we would obscure important

phonological universals in both synchronic and diachronic descriptions; sounds do function in classes, classes which are specified by the features common to their members.

Because the question of internal vs. external evidence has assumed such an important role in discussions on psychological reality, I have asked the symposium participants to present their views on this question.

Each participant has also received one or more questions specific to his or her paper. Let me mention these.

Campbell presents some interesting evidence from Finnish and Kekchi showing the reality of certain posited phonological rules and Morpheme Structure Conditions. He discusses language games played by speakers of these languages. The game data support the rules posited by linguists using internal evidence. Suppose in the language games, these rules were not evidenced. Can one conclude, then, that the P-rules, and MSC's do not exist? That is, what does one do about negative evidence?

This, of course, is not simply a problem that is faced by Campbell, but one faced by all linguists, and, in fact, by all scientists.

Cutler also uses "external" evidence, this time from speech errors, to show that "morphological structure is psychologically real in that English speakers are aware of the relations between words and can form new words from old." She also concludes that "The principles underlying lexical stress assignment are psychologically real in the sense that speakers know the stress pattern of regularly formed new words." This, however, she suggests is in keeping with a "weak" version of psychological reality, which claims simply that speakers can draw on their knowledge of the grammar, as opposed to the "strong" version which would claim that the rules are isomorphic to processes.

It would be interesting to know what kind of evidence would be needed to support the strong version of psychological reality in relation to the posited stress rules of English. What, if anything, does the following error tell us about the psychological reality of the nuclear stress rule?

(Note: for those readers who are not fans of American basketball, Jim West was a famous basketball player with the

Los Angeles Lakers. The meaning of the phrases is paraphrased.)

(1) Target: Jim West Night Game. (The game to be played for the special occasion called Jim West Night.)

Error: Jim West Night game? (the night game played by Jim West.)

Derwing, in his preprinted paper as well as in other of his published works, seems to reject a concept which I hold, i.e. the difference between linguistic knowledge and linguistic behavior. I am therefore interested in how he can find support for psychologically real grammars or rules, given the great variation, including speech errors, false starts, ungrammatical sentences, neologisms, even sounds not ordinarily found in the language that one finds among different speakers of the same language, and even within one speaker on different occasions in both speech production and perception. Is it possible to find exceptionless regularities in behavioral data which permit any generalizations at all? Suppose, for example, one finds five speakers who, to use one of Derwing's examples, relate fable and fabulous, and five who do not. Can we conclude anything? Or should we be constructing individual grammars for each speaker at a single point in time? Or can we conclude instead that, since even one speaker draws certain generalities, the rules which represent them must be psychologically real and permitted by the theory of phonology?

Dressler has distinguished between "naturalness", "productivity" and "psychological reality". How do they relate? Is it possible for a phonological rule to be psychologically real but highly unproductive? And how would such a rule manifest itself. Is there some way that these aspects of language should be delineated in a theory of grammar?

Gussman differs from some of the earlier papers in pointing out that we can not depend on external evidence in our attempts to validate or test phonological hypotheses because it is often the case that different kinds of external evidence are contradictory. It is therefore of interest to know what kinds of constraints he believes should be placed on grammars and how we can

find evidence in support of these constraints. Even while he argues that external evidence may be unreliable, he provides such evidence to argue for phonological representations which some linguists would call "abstract". Is this in itself contradictory?

Hale presents a principle which he suggests is needed in a theory of language, the recoverability principle. How is "recoverability" related to psychological reality? Since the principle refers to an evaluation metric for grammars, i.e. a measure by which we can compare the value of grammars, can the metric itself be used to judge whether a grammar is psychologically real? Or, perhaps even more important, how do we judge the psychological reality of any proposed evaluation metric?

Linell gives us a number of interesting definitions. He defines phonology as "language specific phonetics" and rules as "norms". It is thus not immediately clear what the contents of a theory of phonology as distinct from a theory of phonetics would be.

Finally, Skousen has argued that a linguistic description must be directly inducible from the data. At the beginning of this paper I quoted a statement from Jacob which strongly contradicts such a view. The particular paragraph I referred to ends with a further statement: "[Scientific advance] always involves a certain conception about the unknown, that is, about what lies beyond that which one has logical or experimental reasons to believe." Certainly a linguistic description, in the form of a grammar, should be a "scientific advance", an hypothesis, a theory, which goes beyond the collected data. If Jacob is right, why should stronger or different requirements be placed on linguists than are placed on other scientists? And is it possible for us to discover "new truths", to make "new advances" if we are forced to induce all our hypotheses directly from the data?

These are the questions that have been posed for the panelists. We are sure that there are many other questions from the audience which we look forward to hearing.

Whatever our disagreements, we who are the participants of this symposium agree, as I am sure all in the room agree, that

to whatever extent possible we are seeking the "truth", we are seeking a theory of language, and in particular a theory of the sound systems of language, which will bring us a little closer to understanding the beauty as well as complexity of the abilities of the human mind.

#### References

- Chomsky, N. and M. Halle (1965) "Some controversial questions in phonological theory", *Journal of Linguistics* 1, 97-138.
- Hockett, Charles F. (1942) "A system of descriptive phonology", *Language* 20, 181-205.
- Jacob, F. (1977) "Evolution and Tinkering", *Science*, 196.4295 June 10, 1161-1166.
- Klatt, Dennis (1979) "Lexical representations for speech production", paper presented at the International Symposium on the Cognitive Representation of Speech, Edinburgh, July 29 - August 1, 1979.
- Perrin, J. (1914) *Les Atomes*, Paris: Alcan.
- Schachter, P. and V. Fromkin (1968) "A Phonology of Akan: Akwapem, Asante and Fante", *Working Papers in Phonetics*, No. 9, University of California, Los Angeles.
- Zwicky, A. (1975) "The strategy of generative phonology", in *Phonologia* 1972, Dressler and Mareš (eds.) 151-168.

#### COMMENTS FROM THE PANELISTS

L. Campbell stated his acceptance of the generative phonology goals of descriptive adequacy for particular grammars (which means we should aim at psychologically real grammars) and explanatory adequacy for theories. This requires evidence as to what psychological reality is. Campbell claimed that we cannot find the answer on the basis of internal evidence alone, and one must give greater relative weight to the importance of external evidence. He stated his concept of psychological reality: what is in the head of speakers, i.e. the traditional definition of competence. The more interesting question, he said, is not what psychological reality is, but how do we find out what it is, suggesting that this can only be accomplished by the use of external evidence.

Campbell's answer to the question concerning negative evidence was a simple one: if there is no evidence, there is no evidence. We can conclude nothing. He suggested that a more interesting question concerns counter evidence, which must be used to invalidate theories. He denied the existence of conflicting evidence, despite the reference to such by others. Rather, he suggested that such seeming contradictions are the result of wrong interpretation, theory, or practice.

A. Cutler stated that as she was the lone psychologist on the panel, she would emphasize the "cognitive reality" part of the symposium title by citing some psycholinguistic evidence that prosodic structure is psychologically real. She supported and illustrated her notion of psychological reality by reference to the temporal structure of English, which language is said to exhibit a tendency towards isochrony, in that speakers adjust the duration of unstressed syllables so that stressed syllables occur at roughly equal intervals. She pointed out that there is, however, little evidence that English is physically isochronous; the psychological reality of isochrony is much stronger.

Firstly, English speakers certainly perceive their language as isochronous. In a recent study Donovan and Darwin (1979) presented listeners with sentences in which all stressed syllables began with the same sound, e.g. /t/, and asked them to adjust a sequence of noise bursts to coincide temporally with the /t/ sounds in the sentence. They could hear both sentence and burst sequence as often as they liked, but not together. Donovan and Darwin found that the noise bursts were always adjusted so that the intervals between them were more nearly equal than the intervals between the stressed syllables in the actual sentence--i.e., the listeners heard the sentences as more isochronous than they really were.

Secondly, there is the role of rhythm in syntactic disambiguation. Lehiste (1977) argues that speakers trade on listener expectations by breaking the rhythm of utterances to signify the presence of a syntactic boundary. Durational cues certainly seem to be the most effective at resolving syntactic ambiguities (see, e.g., Streeter, 1978); and recent work by Scott (forthcoming) has demonstrated that boundaries are indicated not merely by

a pause or by phrase-final syllabic lengthening, but crucially by the rhythm--the fact that the foot (inter-stress interval) containing the boundary is lengthened with respect to the other feet in the utterance. Moreover, in a further study of syntactically ambiguous sentences (Cutler & Isard, in press), it was found that speakers tended to lengthen the foot containing the boundary to an integral multiple of the length of the other feet, i.e. "skip a beat" and thus maintain the rhythm.

Finally, there is relevant speech error evidence (Cutler, in press): when an error alters the rhythm of an utterance (a syllable is dropped or added, or stress shifts to a different syllable), it is almost always the case that the error has a more regular rhythm than the intended utterance would have had. In the following examples (syllable omission and stress error), each foot (marked by /) begins with a stressed syllable:

(1) /opering /out of a /front room in /Walthamstow

(Target: /operating /out of a /front room in /Walthamstow)

(2) We /do think in /specific /terms

(Target: We /do think in spe/cific /terms)

The number of unstressed syllables between the stressed syllables is more equal in the errors than in the target utterances. The consistent pattern of such errors supports the notion that isochrony in English is psychologically real: the speakers have adjusted the rhythm of their utterances to what they feel it ought to be.

B. Derwing began his discussion agreeing with Popper (1965) who stresses the importance of the testability of a theory. He then discussed a view which he characterized as that of "autonomous linguistics". According to Derwing, this view holds that there is or may be an idealized natural language system which can be scientifically investigated apart from considerations of the minds and bodies of individual language users. In arguing against such a position, he said that its origins can be traced to a philological notion that a language is an organism complete unto itself and subject to its own unique laws of evolution and change. He referred to a statement of Jespersen that the essence of language is human activity between a speaker and a hearer, and that

these two individuals should never be lost sight of if we want to understand the nature of language and of grammar. Jespersen wrote that words and forms were often treated as if they were things or natural objects with an existence of their own. Derwing agreed that such a view is fundamentally false since words and forms exist only by virtue of having been produced by a human organism. For these reasons, Derwing stated he does not embrace the goal of constructing a theory of language, per se, or a theory of possible grammars.

He suggested that modeling the language user is a better goal, since there can be no doubt that speakers learn something when they learn to speak and understand their language, that they know various things as a consequence of this learning, and that they engage in various kinds of internal activity when they put this knowledge to use. The details of this activity and knowledge are amenable to a wide variety of tests. It is thus not the concept of psychological reality which bothers Derwing, but the concept of autonomous linguistics. In fact, he suggested that the question of psychological reality is debated in linguistics only because there are still a large number of linguists who refuse to admit that linguistics is, or at least should be, a branch of psychology.<sup>1</sup>

Derwing stated that only external evidence can provide definitive answers; such evidence is in fact external only from the standpoint of a theory which ignores it. Both kinds of evidence are useful grist for the same mill.

He concluded by saying that it makes no sense to talk of a true theory of natural language since the object of that investigation probably does not exist. The concept of an idealized, monolithic system of language is a notion we can get along very well without. We can, however, subject claims about human linguistic knowledge and abilities to the test of truth. In this enterprise internal evidence is important and suggestive but hardly conclusive.

---

1) In his remarks Derwing did not cite Chomsky (1968) who may have been the first in recent linguistic circles to consider linguistics as "the particular branch of cognitive psychology".



W. Dressler stated that he conceives of psychological reality in the "weak" sense (Cf. Cutler, vol. II, p. 79-85) in that he is trying to account for the competence of linguistic behaviors. His stated approach is to elaborate a deductive theory of natural phonology and a deductive theory of natural morphology, starting from a few basic theoretical concepts. Conflicts concerning naturalness as pertaining to phonology, morphology, the lexicon, etc. would be derived from the theory. Therefore, hypotheses about the psychological reality of these different types of competence would be derived and tested if the intervening variables in each domain of evidence are controlled.

Dressler stated his disagreement with the Chomsky/Halle (1965) statement quoted by Fromkin in which they say the task for the linguist or the child learning the language is similar; the intervening variables for the two are too different for this to be so. Furthermore, he stated that we should not overemphasize child language acquisition at the expense of other kinds of evidence; it is not the privileged domain, and in fact could lead to wrong conclusions. Besides, massive restructuring of the grammar occurs later.

In Dressler's view, external evidence is not extraneous or some sort of supplementary confirmation or disconfirmation, but a central part of the testing procedure. Thus, external evidence can show that an analysis is wrong. He illustrated this with an example from Italian. The masculine article has two forms, il and lo. Phonological and morphological internal evidence suggest overwhelmingly that lo is the basic form. Yet, an Italian asked to give one form in isolation will produce il. Second, the hesitation form, before pause, is il. Finally, change in progress argues for il. These three kinds of external evidence confirm each other and override the internal evidence. The reason is because the techniques for handling internal evidence have mainly been devised for regular phonological and morphological processes and the system of the Italian articles is neither phonologically nor morphologically regular.

E. Gussmann stated that, if phonological descriptions are to be psychologically real, either in the strong or the weak sense, if, that is, they have some kind of correlates in the mind of the

user, then the basic question is how we can check or verify the reality of the proposed description. He suggested more caution in evaluating external evidence, pointing to the surprising and, in some cases, contradictory results in direct experiments. Specific examples of this are shown in experiments conducted related to the English regular plural formation rule. In some experiments, subjects responded only 50% in the predicted way, but in others 100% of the forms were those predicted by the regular rule. These experiments say little about whether the English plural rule is productive or psychologically real, but do call for a theory of linguistic behavior which can explain the strange results. What needs to be explained is not only why say, 70% of the answers obtained conformed to the predicted regularity, but, also why 30% failed to do so. In other words, he suggested, one cannot conclude there is no regular rule even when one finds that 30% (or more) responses of subjects in an experimental situation are unpredicted by that rule.

This problem relates to the relative roles of internal and external evidence. Internal evidence, he declared, is primary because it is only in reference to such evidence that external evidence makes any sense.

He went on to discuss the need to reconcile external and internal evidence, pointing to the Dressler proposal for representing the velar nasal in German as deriving from /ng/, and the M. Ohala argument in favor of an abstract schwa in Hindi. It is noteworthy, Gussman claimed, that such cases are usually disregarded by proponents of concrete phonology. Given these abstract analyses, supported internally and externally, one should try to formulate the principles speakers must have access to in formulating such rules and representations. Presumably, he added, one would want these principles to be part of a theory of phonology rather than the phonology of a particular language. It is such principles that we should be seeking.

K. Hale addressed the question of his conception of psychological reality, by stating the question can only be answered when related to the linguist's view of the nature of language itself. In his view, language is a complex human capacity, comprising autonomous, but interacting, systems, each of which has



its own inherent principles of organization. Psychological reality, according to such a view of language, is the goal of linguistic inquiry. It is not given a priori. A logical consequence of this is that it is impossible to ask whether a given linguistic analysis is psychologically real or not, independent of the notion of what is the most highly valued grammar. Thus, the psychologically real, or better still, the most real analysis in a particular instance can only be the one that is best according to some appropriate evaluation metric, functioning internal to the particular framework in which a particular analysis is cast and resulting in some natural way from that framework. He added that, in his candid and probably unpopular view, the traditional generative grammarian's notion of a simplicity metric is on the right track. The problem is to have the right metric, no simple matter.

In discussing the question of internal vs. external evidence, he said he finds it difficult to make the distinction, preferring to distinguish between good and bad evidence. When a field linguist is faced with two or more possible analyses of some data, (s)he needs to look at any kind of evidence to decide. In the case of the Maori passive which he discussed in his paper (vol. II, p. 108-113), the analysis he arrived at after looking at ten different kinds of evidence was the unexpected one, setting up a conjugation system among verbs rather than presenting a purely phonological analysis. Yet the phonological rule analysis would probably be the one required of any student who wanted to pass a phonology course. Hale argued that strictly linguistic reasons favor the morphological analysis, referring to Jonathan Kaye's "recoverability principle". This principle also appears to operate in Papago, to select an analysis which could be considered to be just the opposite from that in Maori, although the surface phenomena are identical. This principle may then be a subcase of a more general simplicity metric, affirming the importance of such linguistic principles. He concluded by stating that the psychologically most real analysis will be that most highly valued by a valid simplicity metric.

P. Linell argued for a behavioral performance perspective on language, stating that a language should be viewed as a system

of grammatical and phonological phonetic conditions placed on the stream of meaningful and phonetic communicative behavior. He thus would assign a role to phonological form both as related to plans for the pronunciation of the expressions in question and as related to perceptual schema. Phonological entities are phonetic entities, i.e. phonetic behavioral articulatory plans, intentions, perceptual schemas etc. There are phonological aspects of morphological formation patterns which he said also belong to other components of the grammar, but these, too, concern surface phonetic entities.

Linell suggested that whether one considers psychological reality a non-issue depends on one's theoretical preference. If a language is seen exclusively as a set of abstract sound-meaning correspondences, isolated from behavior and communication, it probably is. Thus, he maintained, autonomous linguistics aims at capturing all detectable generalizations at all levels, and this is a legitimate concern. But if one is interested in psychological reality, Linell proposed that it is necessary to look at production and perception behavior, language learning, and language storage. A language user does not need all the linguists' generalizations and it is thus doubtful that these are psychologically valid. It is more likely, he claimed, that there is great redundancy in the grammar leading to processing short cuts, heuristic routines, parallel strategies etc.

In arguing against formal conditions on rules, or principles, he stated that too often such discussions are pointless since when, for example, we raise the question of recoverability, why should morphophonemic forms be recovered at all, by whom are they supposedly recovered, and for what purpose.

The problem cannot be solved by experimentation, he added, unless we know how to interpret the hypotheses we are testing. If, for example, we find speakers make the vowel substitutions predicted by the vowel shift rule in SPE, we should not conclude that the way the rule is formulated is correct. (Chomsky & Halle, 1968) Or if speakers relate fable and fabulous it is a non-sequitur to conclude that there is one morpheme form underlying both words. This is the generative way of describing the relationship, but there are other possibilities.

Linell concluded with the suggestion that it may be artificial to separate out psychological reality from social and biological reality. What we want is a true synchronic theory of the linguistic practice of language users.

R. Skousen suggested that the psychologically real descriptions which we seek may not be composed of rules such as the kind that have been postulated, or any rules at all. Although linguists may characterize behavior in terms of rules, it is not certain that linguistic behavior itself is rule-governed.

He illustrated his point of view by a discussion of "probabilistic" rules. He considered a hypothetical language in which the verbal past tense is realized by one of two forms, in what has been called in the past free variation. But, suppose in observational studies it is found that a given speaker produces one of these forms two thirds of the time, and the other, one third of the time. He provided reasons why one should not posit a rule which specifies the probability of occurrence of either form in that speaker's grammar. A linguist can construct such a rule, but this does not mean that a speaker can or does construct a rule of this form.

He followed up this example with a discussion on apparent regular rules with exceptions and questioned whether in many of these cases we should conclude that the speaker utilizes a rule rather than looking for specific forms and then using these forms analogically to produce new and novel forms.

#### DISCUSSION

A discussion ensued, participated in by the panelists and by the following speakers from the audience: C.J. Bailey, R.P. Botha, J. Bybee Hooper, R. Coates, T. Gamkrelidze, W. Labov, A. Liberman, L. Menn, J. Ohala, and J. Ringen. There will be no attempt to cover all the interesting points presented.

A number of the discussants continued on the topic of internal vs. external evidence. Ohala posited that this is a false dichotomy, a point made earlier by Hale, since evidence is evidence. He suggested, however, that there is a continuum in the quality of evidence, since some evidence may be less ambiguous and more capable of refinement than other evidence. He

stated that "internal evidence" is highly ambiguous as to what it reveals about psychological entities; evidence from speech errors is of slightly higher causality, and evidence from experiments the least ambiguous and the most capable of refinement because of experimental controls.

On the same question, Bybee Hooper referred to the external evidence used to support the velar nasal as deriving from /ng/ and said that there are other interpretations which can be made, thus warning against making unwarranted assumptions about linguistic structure from such evidence. Both Gussman and Campbell agreed that unwarranted assumptions shouldn't be made about anything.

Hale pointed to the possibility that there may be opposing analyses for which no external evidence is available, and suggested that it is highly possible that a child confronted with a language has a problem similar to that of the field linguist who has only the language data. He suggested that we therefore need some internal principles which permit both the linguist and the child to come up with an analysis. He pointed to problems in interpreting external evidence like that of language games. He has found that in Australia, where secret languages are elaborate and a key intellectual activity among the aboriginal people, some are very good at these games and others very bad. Thus one gets variable data.

Labov followed the lead of Linell's suggestion that one must consider other forms of reality such as social reality, and, in fact, argued that this may have greater importance than psychological reality. He pointed to evidence from child language acquisition showing that children use different strategies before their grammars converge, and he said such differences probably persist in the more irregular portions of the language for some time. In his study of Philadelphian English, he has found that some Philadelphians use a complex rule to derive two phonetic vowels, whereas for others, it appears, two underlying forms exist. Much of the evidence we seek refers to the social reality of the system rather than the processing of individuals.

Bailey also considered the importance of language change, going so far as to say a dynamic approach must be used rather

than a static one in looking at language.

Campbell also added to the discussion on social factors by pointing to the fact that they can complicate phonological descriptions. He has found that in some societies the avoidance of "dirty words" causes phonological complications. Dressler noted that considerations of social reality and the social and communicative function of language was key to a concern for universals in phonology. In discussing variation across individuals Derwing noted that sociological reality was nothing more than a sum of the psychological reality of many individuals. If, he said, we are studying language users, we do not expect them to be the same. Linell suggested that rules should be construed as socially acquired and socially shared, which, he added, is the traditional notion of a rule as a norm for behavior.

Ringen and Botha both discussed the role of the philosophy of science in theory construction and validation. Botha stated there is no such thing as the problem of psychological reality of phonological descriptions. There may be a problem, and this depends first, on the aims of the theory, and second, on the philosophical approach of the linguistic scientist. The notions of "truth", "reality", and "evidence" are theory bound. Ringen also noted the relevance of philosophical questions. He also affirmed the importance of theories of performance in deciding whether evidence is internal or external.

Cutler also argued for the need for a theory of performance but, as a psychologist, pointed to the difficulties in attempting to set up psychological experiments which would get at the strong version of psychological reality. Coates also stressed the importance of working with psychologists in our attempts to establish the kinds of association between linguistic units which exist. The notion of units was discussed by Lieberman, who stated that the basic task for phonology is to segment the non-discrete speech signal into the correct discrete segments.

Gamkrelidze noted that the goal of constructing a theory which would provide for psychologically real grammars was not one which arose with the transformational linguists, who, instead, he believes placed their emphasis on cybernetic considerations. He pointed to the difficulties, however, of trying to

determine what is in the mind of speakers, from their utterances, which parallels the difficulty of trying to determine the inner mechanisms of a clock from watching the hands move. Many models can be constructed which give the same output but only one model is the correct one. This point was similar to one made by Skousen in discussing the need for real world interpretations of formal linguistic constructs, providing an interesting analogy with a formal system of Euclidian geometry which can only have "reality" when the formal primitives are given substantive interpretations.

Menn was concerned with the fact that linguists, or some linguists, seem to ignore the variety of things which can legitimately be considered knowledge and the necessity of distinguishing among them. SPE ignores the degree of rule productivity, she noted, and most experimental linguists ignore the difference between active and passive knowledge and the difference between explicit metalinguistic knowledge ("I can tell you that word A contains morpheme B") and implicit knowledge ("I guess that word A is more likely to mean something about rocks than sugar.") We need to set up sufficiently subtle experiments to be able to differentiate between these phenomena, she said.

To conclude the symposium, the moderator, Fromkin, presented some of her own thoughts. She agreed that it is not possible to proceed without any biases or a specific philosophy of science. One would hope, however, that despite different philosophies, linguists will provide increasing information which will reveal something about the phonological systems of the languages of the world.

She referred to some of the arguments concerning "autonomous linguistics" and expressed confusion as to what that phrase really does mean, or why some people consider it negatively. No one can deny that language is used in society, that language is a product of evolution, that there are brain mechanisms underlying language, that language is used by speakers in producing utterances and in comprehending speech, that it is used for humor, for making love, for expressing hate, for selling soap, but, she asked, why is it not legitimate to attempt to study the language systems which underlie all these uses, to investigate language

per se. The history of science shows the isolation of different facets of reality in order to better understand them. Do we need to study the persuasive and disgraceful use of ambiguities by advertising agencies before concluding that for some speakers of English writer and rider are homophonous even though write and ride are not? And that the homophony arises from an "alveolar flap rule"? Whether or not one believes in the reality of rules, in describing the sound patterns of English, we certainly must reveal this "fact".

This does not mean, she added, that we can ignore the bridges between one part of the complex phenomena and another. But it certainly is legitimate to say that human language exists and we should try to understand it. The question then arises as to whether language is a cognitive system which can be viewed apart from the behaviors of those who have acquired it. Those who hold this opinion point to various kinds of evidence to support it. For example, many if not all of us produce utterances which we, in hearing a tape of our own speech, will regard as "improper" or ungrammatical. This judgment must come from some stored knowledge. Clearly we can and do say, produce, and understand the meaning of utterances that we also declare to be ungrammatical sentences. Thus utterance is not equal to the theoretical construct, sentence.

Fromkin continued her discussion on "autonomous linguistics" saying that the pursuit of language per se may be a worthy one. This does not imply that linguistics is not a subset of psychology. Derwing's dichotomy does not necessarily hold, if we view language as a system of knowledge that is a mental reality. There are of course many subsets of psychology. One can pursue research in the field of vision without conducting research on auditory perception. Furthermore, psychology is concerned with behavior but not exclusively so. There are as many differences of opinion among psychologists as there are among linguists, many stemming from differing philosophical views. Fromkin stated that she could probably point to as many psychologists who agree with her view of the aims and proper subject matter of linguistics as can Derwing in support of his views.

However, she wished to emphasize that this does not mean that the construction of performance models is not a worthy one for linguists. Her own research has been primarily concerned with performance, but she added that this research has been guided by the insights provided by linguists working on language structure, rules, and representations.

Failure to distinguish between linguistic behavior and knowledge would create problems for those analyzing speech errors. Similarly, the study of aphasia shows that in many cases the linguistic deficits are performance deficits, while the stored grammar is intact. Otherwise one could not explain why an aphasic patient is capable of production, retrieval, and perception on one day, and incapable of one or the other aspect of performance on another occasion. Manfred Bierwisch pointed to this discrepancy many years ago when he posited that most aphasia symptoms can only be explained as performance breakdown.

Fromkin concluded with a quote from Poincaré (as cited in Chandrasekhar, 1979):

"The scientist does not study nature (only) because it is useful to do so. He studies it because he takes pleasure in it because it is beautiful. If nature were not beautiful it would not be worth knowing and life would not be worth living."

She ended by saying that we who are interested in human language know how meaningful this quote is, since human language, like all of nature, is beautiful, and the study of it is therefore worth doing.

#### References

- Chandrasekhar, S. (1979): "Beauty and the quest for beauty in science", Physics Today, July 1979, 25-30.
- Chomsky, N. (1968): Language and mind, New York: Harcourt Brace Jovanovich, Inc.
- Chomsky, N. and M. Halle (1965): "Some controversial questions in phonological theory", Journal of Linguistics 1, 97-138.
- Chomsky, N. and M. Halle (1968): The sound pattern of English, New York: Harper & Row.
- Cutler, A. (in press): "Syllable omission errors and isochrony", in Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler, H.W. Dechert (ed.), The Hague: Mouton.

- Cutler, A. and S.D. Isard (in press): "The production of prosody", in Language production, B. Butterworth (ed.), London: Academic Press.
- Donovan, A. and C.J. Darwin (1979): "The perceived rhythm of speech", Proc.Phon.9, vol. II, 268-274, Copenhagen: Institute of Phonetics.
- Lehiste, I. (1977): "Isochrony reconsidered", JPh 5, 253-263.
- Popper, K.R. (1965): Conjectures and refutations, New York: Basic Books.
- Scott, D. (forthcoming): Perception of phrase boundaries, Ph.D. Thesis, University of Sussex.
- Streeter, L.A. (1978): "Acoustic determinants of phrase boundary perception", JASA 64, 1582-1592.

## SYMPOSIUM NO. 4: SOCIAL FACTORS IN SOUND CHANGE

(see vol. II, p. 185-237)

Moderator: Einar Haugen

Panelists: Henrik Birnbaum, Ivan Fónagy, William Labov, Jørn Lund,  
Bertil Malmberg, and Fred C.C. Peng

Chairperson: Martin Kloster-Jensen

## EINAR HAUGEN'S INTRODUCTION

1. The Contributors and their Papers. Each of the invited speakers in this symposium has done research and thought deeply about the topic of linguistic change. They range from newcomers like Lars Brink and Jørn Lund to elder statesmen like Bertil Malmberg. It is one of the prime purposes of such congresses as this to bring together representatives of different views, different ages, and different countries, so that their ideas may be discussed face to face. Unfortunately, each contributor is limited by the format of the occasion to a short presentation in print of the main results of his research and an even shorter presentation by word of mouth. My function as moderator has been the pleasant one of summing them up and showing how together they constitute an advance toward our understanding of the central problem that is the topic of this symposium. One difficulty is that the authors deal with many situations that I do not know firsthand, and that they take up different aspects of the problem itself. In some cases I have had to go back to other work by the same and other authors to clarify the problem in my own mind.

2. Theorizers and Empiricists. The contributors fall into two categories, which I shall call "theorizers" and "empiricists". The "theorizers" are those who base their discussion largely on informal observation from which they make more or less intuitive generalizations. This is not a pejorative description, for in this field I count also myself. I would count among them Birnbaum, Fónagy, and Malmberg. The others are "empiricists" because they present actual field work, much of which has been statistically treated, so that their conclusions give the refreshing impression that we may be able to treat an old problem in a new way, namely by direct observation. I find this approach most exciting, since it builds on forms of data gathering that have become possible once we had tape recorders, computers, and spectrographs. Phonetic

change used to be considered as something we could observe only over centuries. We are now told that we can catch it on the wing. Instead of observing its results only, we can now see it going on. This development appears especially in the papers of Brink and Lund, Labov, and Peng. It has made possible an empirical sociolinguistics, of which earlier investigators could only dream.

3. The topics. I shall first present a very brief statement of the contents of each paper, beginning with the theorists. Birnbaum is largely concerned with criticizing a linguistic model of decoding advanced by Henning Andersen under the name of "abductive". He does not believe that it can account for the rise of innovations in a homogeneous speech community, a construct which in any case he rejects. Fónagy is here concerned primarily with intonation and its historical development. He rejects all notions that it is a "universal" or that it is a fixed, non-arbitrary and motivated phenomenon. Malmberg sees a "state of language" as "a harmonious achronic system or rather complex of systems" within which the speaker may choose according to situation. His chief example, which he has previously studied in detail, is the Parisian vowel system, or rather its "maximum" and "minimum" systems. He regards the rise of "minimum" systems as the result of a "simplification" that is typical of persons living on the social and spatial periphery of a society. Brink/Lund (as I shall call them jointly) have gathered a vast amount of data on the phonetics of Copenhagen speakers born between 1840 and 1955, fully presented in their massive two-volume *Dansk Rigsmål* (Copenhagen, 1975), unfortunately available only in Danish. Basing themselves primarily on phonograph recordings going as far back as 1913 as well as whatever printed materials are available, they have identified up to sixty regular phonetic changes. They have divided their speakers into two social groups, speakers of "high" and "low" Copenhagen. Labov's work has dealt with a variety of American groups, beginning in the island of Martha's Vineyard in Massachusetts, continuing on New York's lower East Side, and currently in Philadelphia. He has concentrated on Black youth, but has worked with all colors and social classes. Finally, Peng bases himself on extensive data gathering in Tsuruoka, Japan, by his colleague Nomoto. This was a sample first drawn in 1950 and then reexamined in 1971. The novelty in his theory is that one generation is sufficient to

identify the process of sound change. Labov's period is in some sense even shorter, since he studies different age groups synchronically and assumes that young people will carry their innovations on into adulthood. We are fortunate in having a wide variety of data bases, from three continents, as well as considerable variety in theoretical approaches.

4. Stability and Change. Except in immigrant communities, every community studied so far has enough stability of language so that each generation can communicate with every other. At the same time language is known to be changing at a rate such that after some unspecified number of generations it will become unintelligible to its ancestors. These basic facts determine the possibility of two complementary views: that language is stable and can form the object of synchronic study, and that language is constantly changing so that it can form the object of diachronic study. In their extreme form both views become unrealistic, e.g. in assuming complete homogeneity or complete fluidity. Members of the Prague School (e.g. Havránek, see Garvin 1959) described "elastic stability" as desirable in a standard language, but in fact they were only defining the nature of all language, "standard" or not. Labov has invented the latest synonym for this term in his "orderly heterogeneity", which is as much a construct as Chomsky's "ideal homogeneity" to which he opposes it. Both agree that language is "structured", i.e. amenable to description by rules. Chomsky's are categorial, Labov's variable, but there is structure in both. The step from categorial to variable rules is a great step forward in descriptive linguistics, but it was foreseen in historical linguistics, and especially in dialect geography.

Here it is useful to emphasize the concept of "choice" as used in Malmberg's paper. Variable or conflicting rules mean that individuals have the freedom to change language within wider or narrower limits of acceptability. But none of these rules are very helpful so far in predicting the future. Any attempt to predict sound change has to face the problem of showing why people make decisions as they do. But this involves going back into their individual and collective psyches to study their unconscious motivations, an infinite regression that leads us far outside the realm of most linguists' competence, though some have loved to



speculate about it. A careful study of the tiny rule changes in Copenhagen speech pinpointed by Brink/Lund suggests that at any given moment in time there is an enormous amount of unstructured heterogeneity, of vacillation and uncertainty. This may either continue, or be resolved by a later generation, and it may lead either to innovation or to regression.

5. The Problem of Actuation. It is hardly surprising that living language abounds in heterogeneity. It is more surprising that there is no more of it than there is. The basic reason for heterogeneity has been evident ever since men stopped believing in such myths as the Tower of Babel. Recent linguists have re-discovered the fact that language is innate and universal, but the most universal fact about languages in the plural and concrete is that every one of them has to be learned anew by every human being born on this planet. He or she is born to human parents and in a human society, surrounded by the speech output around it. That output becomes the input to the child's own processing of the language for reception and eventually production. The study of the child's language learning (which for some arcane reason has come to be known as "acquisition" -- perhaps it is part of our acquisitive civilization) has become an important field of research. We may look to its results for new light on the extent to which the fully formed child's language differs from that of its environment. We do know that eventually all non-defective children learn to communicate in whatever language variety is spoken around them, in spite of the inevitable differences among individuals in talent, appearance, industry, and success. But human beings are not robots and no given language is imprinted by instinct. Try as they will, people will deviate. Call their deviation a "speech error" or a "creative innovation", as you will; it is the germ of a language change.

6. The Mechanism of Diffusion. Given the fact that more or less random innovations occur, we need to pinpoint the process by which they are spread to other speakers. If they fail to spread, they remain speech errors; if they do spread, they become linguistic changes. On this point our symposium speakers show a clear difference. Brink and Lund appear to believe that the innovations are made in childhood and are then retained for life, unless of course the speaker moves into a new linguistic environ-

ment. Their basis for this claim is the recordings they have studied of the same speakers at various periods of their lives. It must be noted, however, that age 15 was the lowest they studied, which is already after the onset of puberty. Many studies have shown, whatever the cause of it may be, that puberty is a period when language tends to fix itself into an adult pattern that most people find difficult to change. Birnbaum emphasizes the importance of the teens as "the age when growing-up speakers, by imitating their elders, attain the same or nearly same pronunciation as their models." He regards such changes as frequently deliberate, and due to fashion within the generation. At the same time he rejects the simple transfer of one generation to another, since there is a "continuous pattern-setting effect of parents on children, teachers on students, leaders on followers, older on younger playmates and fellow workers, more prestigious on less prestigious..."

Against this view Peng entirely rejects the idea that change takes place across generations. He specifically denies Johnson's (1976) view of an accelerating change over three generations. He has found that Nomoto's speakers showed many changes over a period of 21 years. He suggests that while the rate of change may go down as age goes up and reaches a low point around age 35, it never completely stops. He questions Labov's use of "apparent" time studied in synchronically present generations and advocates the use of "real time". Presumably Labov would agree that this is desirable when the investigator lived long enough, or when his informants do, for he (Labov) refers to Hermann's restudy of Gauchat's famous village of Charmey in Switzerland. Peng suggests as an alternative the use of dialect geographical material, with its mapping of horizontal linguistic change. This, too, is a case of apparent time, however, since the dialects exist synchronically, and we can deduce just how or even approximately when the change took place only by the use of comparative-reconstructive methods.

7. Class Correlations. Our speakers also show certain differences of opinion concerning the role played by social and other classes in the actuation of change. Labov has found that in American cities the upper working or lower middle class, that is, the centrally located classes, lead in linguistic change. The speakers who are most advanced are the ones with the highest aspirations for advancement, who also have the largest number of



local contacts outside the community. Malmberg has fixed his view on the central norm of Parisian French and regards simplification as a major factor, which he then attributes to the lower classes and the provincials, who live on the periphery. In Brink/Lund's detailed account of their three-score changes in Copenhagen, however, the role of social class is rather different. To begin with, they deny that there were what we would call class differences prior to 1750. Before that time the speech of Copenhagen was a local dialect like any other, different from its neighbors, but having much in common with them. In the 18th and 19th centuries a class differentiation took place which reduced contact between different strata of society. A distinct lower-class speech developed, which in general was ahead of upper-class speech. Only since 1900, when everyone is sending their children to publicly supported common schools, are the differences leveling out, or in the view of the *élite*, the language is being "vulgarized". Unfortunately, it is difficult to compare Brink/Lund's results directly with Labov's, since they operate with only two classes as against Labov's more refined indices of class membership.

On one point everyone seems to be agreed: that women everywhere are more "refined" than men of the same age or class, i.e. have more features classified as "high". Brink/Lund are not willing to grant the existence of a separate "sexolect", but suggest that women are more sensitive (perhaps rather "sensitized") to social status. Fónagy finds that in Hungarian a final rising intonation has lost its marked value as an indicator of "expressiveness". The reason is that it has now become normal among women and young people.

8. Conclusions. Two of our speakers emphasize that it is not language that changes, but people who change language. Peng writes, "People change, and sound change is simply a manifestation (or symptom) of human change." Malmberg reiterates from his Bucharest paper (1969) that "language does not change; man changes languages." These statements are true, but tautological, unless we are speaking of the adoption of new words or the learning of new languages. Phonology tends to fall below the threshold of consciousness for most speakers, and they are rarely aware of making changes in their own speech. It is only with the greatest caution that we can identify any external social reason for such

unconscious change. Nothing in climate, occupation, physiology, character, or history can be causally connected with such large-scale linguistic changes as the Germanic consonant shift, or *Umlaut*, or the English vowel shift, or even with the decay of inflections in most Romance and Germanic languages.

Brink/Lund even deny that the Copenhagen forms have spread because of the prestige of the capital city. But their claim that they spread "purely by contagion" makes one wonder why they did not spread the other way, during a period when the city was invaded by great numbers of rural immigrants. They believe that new pronunciations spread by virtue of an "inherent plus value", vaguely defined as their being "easier to articulate", and conclude that "sound change is essentially a non-social phenomenon." William Labov, who has done more to correlate social and linguistic variation than anyone else, is equally pessimistic: Bloomfield's assertion of 55 years ago that "the causes of sound change are unknown" is still true.

In spite of the weight of first-hand research and authority which these writers bring to the topic, I cannot let this conclusion stand as the final word of the symposium. I am convinced that the causes are known, but that what is really meant is that the results are unpredictable. Let me briefly sum up my own unsupported and intuitive view of sound change (though it is not unlike that held by Hugo Schuchardt and Otto Jespersen). Sound change is in principle no different from any other change going on in the lives of animate beings everywhere around us. To say that we do not know the causes of change is like saying that we do not know the causes of human fashions, e.g. the length of women's skirts or the shape of men's headgear. We do know that one main cause of human language change is that language is not genetic, but learned, and that no two human beings ever learn anything exactly alike. I do not believe that the parts of any language hang together in Meillet's sense of "tout se tient". If they did, there would neither be sound change nor the development of dialects. I believe instead in what I may call the "amoeba" theory of language, that any aggregation of items we call a "language" or "dialect" is as arbitrary as the movements and splittings of the amoeba. The most important rules of language are simple collocations. Phonetic changes can only have been "actuated" by

individual learners and users, whether as children or adults, who committed errors in hearing or reproduction that were not corrected by themselves or others. Phoneticians can tell us a great deal about the physical and acoustic parameters that favor such errors, but they cannot predict which of them will occur.

To become part of the speech of others, these innovations have to be acceptable to other members of the community. This is the process of diffusion, which has to be both lexical and social. Lexically, the change has to spread from the one item in which it started to other items that in some way are felt to be similar to the first. The neogrammarians' or any other linguistic formulation of such changes or "rules" as they are now called is an ex-post-facto summary of change, not a description of the change itself. As dialect geography clearly shows, a change may stop at any point in its diffusion, before it has spread to the entire lexicon or the entire community. It may even change its domain, be reordered or reorganized, apply to different parts of the system, be lexicalized or grammaticized. "Simplification", which is often resorted to as an explanation, is no real answer, for neighboring dialects fail to simplify in the same way. According to Chen and Wang (1975:267), the final nasal consonant /m/ has been lost in Mandarin, but in Cantonese it is still there. Who could have predicted that? It is vocalized in French, but in English we still have it. A tendency, yes; a universal, no. Besides, in spite of all simplification, every language known seems to be of about equal difficulty, learned at much the same age by children who are exposed to it.

There are too many factors present in every human situation for us to be able to foresee all its possibilities. No sooner has one rule operated for a time than another takes over and messes it up. Such is life, and language is no different.

#### References

- Chen, M.Y. and W.S.-Y. Wang (1975): "Sound change: actuation and implementation", Language 51, 255-282.
- Fónagy, I. (1956): "Über den Verlauf des Lautwandels", Acta Linguistica Academiae Scientiarum Hungaricae (Budapest) 6, 173-278.
- Garvin, P.L. (1959): "The standard language problem -- concepts and methods", Anthropological Linguistics 1, 3, 28-31.

- Jespersen, O. (1933): Linguistica: Selected papers in English, French and German, Copenhagen.
- Johnson, L. (1976): "A rate of change index for language", Language in Society 5, 165-172.
- Labov, W. (1972): Sociolinguistic patterns, Philadelphia.
- Malmberg, B. (1969): "Synchronie et diachronie", Actes du X<sup>e</sup> Congrès International des Linguistes, vol. I (Bucarest), 13-25.
- Schuchardt, H. (1928): Schuchardt-Brevier, Halle.
- Sommerfelt, A. (1968): "Phonetics and sociology", in Manual of phonetics, B. Malmberg (ed.), 488-501, The Hague.

#### COMMENTS FROM THE PANELISTS

Birnbaum did not intend his paper as a major critique of Henning Andersen's abductive model of phonological innovation, for which he has great admiration. He only wished to indicate that it could be improved on some minor points, e.g. the problem of generational sequence. He was concerned with any trend toward excessive schematism. As for being classified as a "theorizer", he wanted to make it clear that he believed in a happy combination of data gathering and theorizing. He agreed that early childhood was the most important period for establishing speech habits, but that puberty also led to readjustments.

Fónagy was stimulated to study French accent after being rebuked for having an 18th century pronunciation on his arrival in France thirty years ago: he made it a habit to place every stress on the last syllable! He has found that French stress is elusive: its placing is a probabilistic function of many variables, including syntax, genre, etc. Today radio and television speakers are increasingly stressing enclitics, which are not stressed in conversational speech.

Labov described his paper as the first report on his Philadelphia study, his largest project so far, using more advanced techniques than his earlier studies. He has adopted the strategy of searching for innovators: where are they in the social spectrum, by sex, class, position etc. How is sound change related to the network of communications and to new ethnic groups that enter society? Can we throw light on change by looking at the people who are doing it? He does not think that the individual is a significant unit: we are dealing with the social pressures which form an individual into a social being as he grows up and assumes

a variety of roles in the social structure. His main motivation in coming to the meeting was to make contact with acoustic phoneticians and the theoreticians who have developed the models we use: Fant, Fujimura, etc. "Ever since 1968 we've made the point that the tools of acoustic phonetics are useful for examining problems of language structure and language change." These tools will require increasing understanding of the mathematical models at the base. A report on the Philadelphia study should be available in three or four months.

Lund, on behalf of himself and Brink, spoke about their findings in the study of Copenhagen pronunciation. They found that "the sound pattern of the single individual will not change significantly after the teenage years unless the linguistic environment is changed rather profoundly." In the book they had taken the position that sound change takes place across generation boundaries, but they did not deny Peng's contention that sound changes in progress can be studied within one generation. But in this case there is often situational variation, with old forms in more formal speech, new forms in more casual speech. Here Malmberg's distinction of maximum and minimum may be applicable, though they found the term "minimum system" problematic. In casual speech there are not only the typical reductions and assimilations, but also subconscious new sound qualities that do not necessarily lead to simplification. Nor can they see anything here in common with aphasic speech or the reduced inventory of phonemes often characteristic of foreigners. They agree with Fónagy that changes in prosody "must be accounted for in the description of linguistic evolution." They question Labov's finding that the most advanced speakers "are those with the highest status in their local community, having found that new pronunciations have low prestige and are often considered vulgar, if noticed at all." They agree with Haugen that most changes are unconscious and that their investigation is difficult to compare with Labov's, since they started from the phonetic variation, and only secondarily examined the social correlation. "No Danish pronunciations are characteristic of the middle classes."

Malmberg noted that his paper "starts from my distinction between a language as a closed, hierarchical system of mutually dependent units, a structure sui generis, in our case the phonemic

system, where any change in the number and/or the relations of these units implies the creation of a new language richer or poorer or differently structured." Further, "a state of language..." (a Saussurean term) "is a sociolinguistic concept which for its full definition needs extra-linguistic parameters." "Every system or subsystem ... can function as one of the layers within a state of language." "The degree of mastery and retention of the complexity... is a question of the strength of the social norms which determine the speakers' behavior. The terms 'maximum' and 'minimum' systems must be understood as abstractions." By "simplification" he referred to phenomena occurring in the social and geographical periphery of normative centers and areas in contact with other systems on the linguistic border, including the diffusion of languages to new areas through colonization. He did not have in mind peripheric local dialects, which can be very conservative. "My principal point is the existence of layers of varying complexity and of norms of varying strength and the (socially determined) choice between different possibilities." "My intentionally provocative formulation at the Bucharest Congress in 1967 was made to stress the importance of the choice factor and that of social evaluation in phonetic/phonemic change."

Peng called attention to the two basic assumptions in his paper: (1) That language change is a change in behavior. Only by studying changes in language behavior can we discover changes in the code. Once this step is taken, one can observe changes within a single generation, without waiting for two or more demographic generations. (2) A random sample is more representative of human behavior than one that is previously stratified for class. In his work in Tsuruoka the same questionnaire was administered to 137 informants chosen at random and interviewed 21 years apart. In this way it was possible to make use of real rather than apparent time. In plotting the changes over time, one gets a straight line, showing that all age groups were affected.

Labov agreed that people tend to preserve their vernacular and gave the example of a mother and a daughter who differed widely in the pronunciation of the /aw/ diphthong. But he granted that people change their norms and only now realized that Peng had been studying the formal responses to norms and not the vernacular. He himself was looking for un-

reflecting speech, "the most systematic motor-controlled speech." No one has studied syntactic change, which may indeed be individual (cf. study by René Agneau of the progressive in 19th century English, showing that e.g. George Eliot made increasing use of the progressive in the course of a half century.) He expressed admiration for Peng's use of real time, but in his own work he preferred to begin with people in the context of their local community. He agreed with Lund that whenever changes rise to the level of consciousness, speakers tend to reject them.

Birnbaum commented on the moderator's summary. He gave an example of women's speech as different from men's: women tend to use an implosive /h/ in a word like jaha. He agreed that prediction is dangerous, and gave an example from Polish, the replacement of nasality in final vowels by diphthongization. Also that we can ascertain the causes of change, but that we cannot always explain them. He found the summary to be an important paper, by virtue of the moderator's including views of his own, perhaps unduly pessimistic.

Haugen as moderator responded modestly that he found the non-systematic parts of language more interesting than the systematic ones, whose existence he had never denied. He found that only by assuming an arbitrary disjunction between the parts of a system could one explain that they could change independently. One example is the well-known fact that an adult learner can speak a language fluently and with virtually perfect syntax and lexicon without ever mastering the phonetic system.

Peng noted that he had speculated on the causes of change and found many factors and mechanisms. He did not feel that the generation boundaries were primary, but the fact that speakers pass on a different language from the one they themselves learned. Diffusion of the code and diffusion of the people who accept it are two concurrent dimensions of diffusion. He challenged Lund to explain how he arrived at his conclusion of non-change on the part of individuals.

Fónagy mentioned retrospective studies of linguistic change in the 16th-18th centuries. They show that there are enormous differences between sound change and sound change. Some changes are dependent on sex (one reason given for a difference in women's speech at that time was that it was not good form for them to

open their mouths too wide), others are not. Some changes are socially dependent, some are word class dependent, others are not.

Lund replied that they had made spot checks of the same person recorded in the same speech situation many years later.

## DISCUSSION

Simone Elbaz (Paris): "Mon intervention n'est en rien polémique. C'est une mise au point. J'ai le plus grand respect pour tous les grands noms cités, mais je m'étonne de l'absence totale de référence aux travaux d'André Martinet depuis le début de ce Congrès, et même dans l'aperçu de M. Rigault hier, qui cite Jakobson, Saussure, Chomsky en oubliant que la description d'Hauteville (1956) a servi d'exemple à bien des travaux ultérieurs.

Je veux rappeler que Martinet a été l'un des premiers à reconnaître et à étudier les changements linguistiques (cf. Economie des changements phonétiques, 1955); il a toujours dit: "Une langue change parce qu'elle fonctionne".

Récemment, il a cultivé et circonscrit la notion de synchronie dynamique qui, différente de la diachronie conçue comme l'étude et la comparaison de deux états de langue et de la synchronie conçue comme constat d'un état de langue, englobe non seulement l'analyse des variantes dans ce même état de langue, mais encore les prédictions de son évolution.

Cette notion de synchronie dynamique me semble intéressante dans le cadre des discussions de ce matière, c'est pourquoi j'ai voulu la présenter. (cf. Evolution des langues et reconstruction, Paris, PUF, 1975)."

Tore Janson (Stockholm): "Language is not only spoken; it is also heard, and the expectations of the hearer must also be changed. So it is important and possible to study the reactions of the hearers, e.g. in experiments with synthetic stimuli. I have done some experiments and would like to get in touch with people working in this area. The results so far are very interesting."

Lars Brink (Copenhagen): "We have tried to show that the forms of a capital city can be spread purely by contagion, according to what we call 'the Napoleon principle': "The enemy is beaten where he is weakest and is immediately enrolled in the

victor's troops." Of course prestige plays a significant role, but not in spreading new pronunciations. The innovations were never felt to be prestigious. Some innovators may be so, but not their followers, and the innovations would therefore drown in traditional forms.

Henning Andersen (Copenhagen): He called for greater precision in the expression of ideas. He did not think Brink said exactly what he meant when he said that a capital like Copenhagen could spread its forms to the countryside. You do not spread changes. It is the people who change their language to conform to the norms as they perceive them in the capital. He then entered a plea against Haugen's view of language as non-systematic or at least finding the non-systematic parts as more interesting. "We won't understand how more or less stray variation that goes on in speech production at all times may become codified and integrated into a system unless we study it in relation to the systems (or the code) that underlie speech production. Labov's study shows that even minute changes are accessible to some degree of subconscious awareness and confirms that what happens when variations turn into a kind of drift is precisely that what could be stray variation becomes a sort of fashion (and here I subscribe to Haugen's view) and is integrated. If we want to explain how changes can be integrated into one system, but not into another, or how changes can occur in one language but not in another, we need to refer to the systems that the stray variations can be integrated into." He then cited Roman Jakobson's opening statement to the Congress, read by Rischel, to the effect that "there is no gross sound matter in language: everything is formed", etc.

Irmgard Mahnken (Saarbrücken): "The question has been raised of how changes can arise in a homogeneous speech community. There are languages which have not changed for a very long time, and others that have been changing and then have stabilized themselves. At least theoretically we need a model of non-change as well as one of change, especially in the development of literary languages. Very little work is being done on the latter, since the social aspects now being investigated are based on living languages. The question of prestige and of social expression can explain many things now under discussion."

Helmut Lüdtke (Kiel): Sound change is predictable. The question is: how and how far? For example, if we knew Latin but no Romance language and wished to predict in what way a Latin word like clave might change in 2000 years, we could choose from the forms written on the left-hand side of the blackboard. Lüdtke suggested that a limited number of possibilities existed, and one would not choose something like akulavic or que. Sound change moves in an irreversible direction, toward shortening. Lüdtke has a theory which he may explain at the next congress. Sound change is reduction: the allegro forms of today are the lento forms of tomorrow.

Eli Fischer-Jørgensen: "I started changing my language when I was fifty and have continued until now. I spoke a conservative form of standard Danish when I was young, and now I find myself using a pronunciation which is approaching what I consider 'vulgar' Danish. This has happened unconsciously and against my will (but the change appears quite clearly from tape recordings). This is quite contrary to some of the ideas presented here." (J. Lund later commented that this might be due to her having a higher linguistic consciousness than most others.)

Richard Coates (Sussex): One often gets the impression that sound change is either community-internal or due to some catastrophic eruption into the community. Coates wished to point out a third mode which has occurred in the literature recently: a new norm external to the community has been integrated into the linguistic system by the adoption of personas by young children. This is exemplified in the work done by Reed in Edinburgh and recently published in the Trudgill volume of readings. Children who were well grounded in the local dialect were able to adopt pronunciation personas taken from TV personalities, disc jockeys, etc. A well-known boxing commentator's mode of presentation was adopted to describe playground fights by particular children. Here is a new norm, a new vector not due to ordinary situational interaction. It is potentially usable independently of the originally appropriate situation. More than one norm is being sanctioned within the system, highlighting once again the dynamic synchrony which has often been mentioned as a feature of these discussions.

Gilbert Puech (Oullins): [In the absence of a written text, the speaker's French is translated into English.] Puech noted

that changes had here been presented as due to social and geographic stratification across a linguistic community. This view should be complemented by studying the need of a social group for a marker of its identity, a change which concerns the weakest point in its system. Therefore he posed this question to Professor Labov: For Philadelphia modifications have been pointed out as due to the lower middle class. Does this correspond to the emergence of this group as a social category which needs to emphasize its identity more strongly by initiating or accelerating linguistic changes? Is it an active or a passive behavior, a consequence of the existing division?

Pierre Léon (Toronto): "(1) Au sujet de la durée des changements -- question posée par Haugen -- certains changements peuvent être très rapides (cf. Léon: L'accent en tant que métaphore sociolinguistique, French Review, 1974). Les ruraux prolétarisés d'un village du centre de la France ont adopté certains traits de prononciation urbaine (parisienne) et prolétaire (ouvriers de la banlieue parisienne), en moins de 10 ans. (2) Ce changement est ce que Léon appelle le résultat d'une conduite idéologique. La nouvelle articulation des ouvriers du village est ce que Birnbaum nomme ici 'a conceptualized (verbalized) mirror image of mental activity' et Fónagy un processus 'métaphorique'. Faudrait-il dire métonymique? (3) Au sujet de savoir qui est responsable de la variation -- question posée par Haugen, Brink, Lund et Labov, Léon donne des exemples des facteurs de la variation dans son village: jeunes, adultes, hommes, prolétarisés. Dans une enquête sur la standardisation des prononciations dialectales de la France (Léon et Léon, à paraître dans les Actes des Congrès de Miami), les facteurs de la variation se groupent en 2 séries oppositives:

standardisation +	{	jeunes    ≠    vieux	}	+ statu quo dialectal
		citadins    ≠    ruraux		
		mobiles    ≠    sédentaires		
		favorisés    ≠    défavorisés		

Tous les facteurs n'ont pas le même poids. (4) Le concept de l'hétérogénéité ordonnée de Labov se retrouve dans les exemples données par Fónagy et se confirment dans les résultats de l'enquête de P. Léon et M. Léon, qui montrent, à côté de la disparition des

systèmes de marques dialectales, une diversification au niveau des types de discours. (5) Le concept de sociolinguistique, tel qu'il est employé actuellement n'est-il pas trop restreint aux phénomènes d'indexation des classes sociales, éventuellement aux catégories sexe et âge? Ne faudrait-il pas tenir compte des marqueurs professionnels (Fónagy) et stylistiques dans une approche phono-stylistique plus large (Léon 1971, Essais de Phonostylistique, Didier) tenant compte des facteurs expressifs des situations de communication?"

Anatoly Liberman (Minnesota): On the predictability of sound change he agreed with Haugen: it can always be explained afterward. There are so many things that can happen that given our framework to-day, the framework of system, which is such a very nebulous thing, we can hardly predict what will happen. Also, some things are more probable than others; but given a proto-language and 100 dialects, it is humanly impossible to predict the future. We can only sometimes predict the past, i.e. explain what has happened, but even that is tremendously difficult.

Birnbaum: "I share fully Professor Elbaz's surprise that in all these papers the name of André Martinet was never mentioned". "In a side comment I referred to Martinet's dictum: 'Language is a balanced system with continuous functional redistribution'. To T. Jansson Birnbaum remarked that we all agree that speech perception is important in sound change. Henning Andersen's whole model is related primarily to perception. To I. Mahnken: "Andersen's model was developed to account for historical changes in a Czech dialect." To H. Lüdtke: "I would not call your procedure 'prediction', but educated guesses about probabilities." Reduction is important, but the factor that counters it is the need of explicitness. These forces are constantly in conflict, and it is very difficult to say which will win.

Labov: (1) On women's speech: we do not all agree that it is more advanced. Where women play a part in national life, they are more sensitive to the national prestige, once a sound change has reached maturity and is stereotyped. They are also normally the leaders in linguistic changes from below or unconscious change, where we are hypothesizing a different kind of prestige. (2) This has not been a panel dealing with restraints on linguistic change. However, following Weinreich's paradigm, many



of the sound changes discussed here do show very powerful unidirectional principles, such as the fact that tense vowels always rise. -- On the question of the upper working class: that is not a final characterization of the group involved because it turns out that the role of these innovators in linguistic change is characterized even better by factors having to do with communications research. They are leaders in certain community networks which are very intense locally, but which reach outside the community, and so we get a relatively homogeneous city dialect. Do they emerge as a new group with a need for identification? "I suspect that Professor Puech's characterization was correct. It is not necessarily a new group. It may be an old group that needs to reinforce its identity. These mysterious factors of prestige which we cannot make explicit may be the result of pressures from new groups entering the community. These are challenging the position of the old group. Just as an adolescent must reassert his position in his parents' community, so the Irish or the Italians or the upper working class may be under pressure from Blacks, Puerto Ricans, and other new groups entering the community. Yes, I suspect that the pressure to reassert identity is the driving force behind this continual renewal of sound change."

Suzanne Romaine (Birmingham): Labov's research is an important attempt to deal with the problem of the transmission of change. But the value of the work being done on social factors in sound change is not (as Labov seems to think) to provide explanations of why language changes, but to give us a taxonomy of how social factors interact with linguistic structure in the implementation of language change.

Haugen: "I think we are still in the midst of a very important and very interesting discussion. I thank you for listening to this segment of a discussion that I am sure will go on at future congresses as well as between congresses."

## SYMPOSIUM NO. 5: TEMPORAL RELATIONS WITHIN SPEECH UNITS

(see vol. II, p. 241-311)

Moderator: Ilse Lehiste

Panelists: George D. Allen, Robert Bannert, Christopher J. Darwin,  
Hiroya Fujisaki, Björn Granström, Dennis H. Klatt, and  
Sieb G. Nooteboom

Chairperson: Claes-Christian Elert

## ILSE LEHISTE'S INTRODUCTION

The title of the symposium leaves open the question of the type and size of the speech units. The contributors to the symposium have indeed chosen to address themselves to units of quite different types and sizes. Likewise, they have approached the problems connected with the temporal structure of speech units both from the perspective of speech production and from that of speech perception. The contributions include highly theoretical papers, papers presenting detailed results of experiments, and papers falling between these two poles. Some systematization appears to be in order. I would like to present herewith a framework within which I believe the issues can be profitably formulated for the discussions which I hope will follow.

The framework involves three dimensions. One of them concerns the relationship between timing control in production and the role of timing in perception. The second dimension deals with the direction of determination in the temporal organization of spoken language: specifically, with the question whether the timing of an utterance is determined by its syntax, or whether there exist rhythmic principles in production and perception that are at least partly independent of syntax. The third dimension follows directly from the previous two and relates to the type and size of speech units. What is the nature of those units, and are they to be established on the basis of a morphosyntactic analysis of the sentence, or on some kinds of independent phonetic criteria?

Clearly both production and perception are involved in oral communication by spoken language, and it would seem unnecessary to elaborate the point. However, I have had occasion to argue--against considerable weight of opinion--that durational differences in production, be they ever so significant statistically, cannot play a linguistically significant role if they are so small as to



be below the perceptual threshold. It would be wise, I think, to remind oneself periodically of "the evident fact that we speak in order to be heard in order to be understood" (Jakobson et al. 1952). I hope, therefore, that in our discussion of temporal relations within speech units, models of production and models of perception will be related to each other.

The second and third questions concern the direction of determination: does phonology follow syntax, or are we dealing with interacting, but parallel hierarchies? Some researchers have developed programs for generating the temporal structure of a sentence on the basis of segments and syntactic structure, without paying any attention to rhythm. This is, I believe, due to a particular theoretical orientation. Generative phonology operates with segmental features; even suprasegmental features are attached to segments. And in a generative grammar, phonetic output is the last step in the generation of a sentence. An independent rhythm component simply has no place in the theory. For these scholars, then, the speech units are segments, phrases, clauses, and sentences. (And it is quite interesting to see them struggle with units not foreseen in the theory, like syllables and phonetic words.) Researchers who are not fully committed to this theoretical viewpoint operate with certain other units, such as speech measures or metric feet. Again, the reality of both kinds of units can be studied from the point of view of production as well as from that of perception.

Practically all the issues I have outlined are treated in the papers contributed to this symposium. Production is the main concern of the papers of Allen, Bannert, Klatt, and Öhman et al.; perception is the focus in the papers of Carlson et al., Donovan and Darwin, Fujisaki and Higuchi, Huggins, and Nooteboom.

In my brief summary of the papers, I shall address some specific questions to the authors, and raise some general questions that I hope will be discussed at the end of the presentations.

Among the papers dealing with production, Bannert considers the relationship between the durations of vowels and consonants in stressed syllables of disyllabic words in Central Swedish--words of the types stöka (V:C) vs. stöcka (VC:). When sentence accent is added to these words, both segments are lengthened, but by unequal amounts. The increase is largest for the long segment of each type of sequence, i.e. the long vowel in stöka and the long

consonant in stöcka. Bannert finds that the temporal structure of quantity is best described by using the concept of vowel-to-sequence ratio,  $V/(V + C)$ , and he proposes that the VC sequences be viewed as units of production and perception.

I have a comment and a question. The comment relates to the observation that lengthening affects the long segment of the VC sequence. It might be useful to recall here that already Trubetzkoy defined the difference between long and short phonemes in terms of stretchability: tokens of long phonemes are stretchable, while short ones are not. Knowing that it is the long element that is stretchable, one could have predicted Bannert's result: that the addition of sentence accent to quantity increases the temporal distance between the two word types.

The question concerns Bannert's proposal that VC sequences be viewed as units of production and perception. I would like to know how such units relate to already well established units such as syllables. Presumably the syllable boundary falls before the single intervocalic consonant in words like stöka and within the long intervocalic consonant in words like stöcka. I find it difficult to conceptualize the psychological reality of the VC sequence as distinct from segments on the one hand and syllables on the other. It seems to consist of non-comparable parts of the two syllables. Where would these VC sequences fit in a hierarchy of units of production? And what is the evidence for the claim that they also constitute units of perception?

The paper by Klatt presents a detailed scheme for the synthesis by rule of segmental durations in English sentences. It is an almost pure example of that approach that starts from an abstract linguistic description and ends up as a sequence of segments whose durations are conditioned by other segments and by syntactic constraints. The paper does not address itself to the question of overall speech rhythm. A companion paper by Carlson, Granström and Klatt is devoted to testing the output of Klatt's synthesis algorithm. Among the interesting results are the observations that certain aspects of the durational pattern are of greater perceptual importance than others. Vowel duration is more important than consonant duration; the durations between stressed vowel onsets seem to constitute a particularly important aspect of sentence structure. Now it is known that English is a stress-timed

language; there exists an extensive literature dealing with isochrony in English, and some of the arguments in favor of the existence of isochrony are quite persuasive. I would like to address a question to the three authors of the two papers, concerning the role of rhythm in the production and perception of English sentences. Would it not be advisable to include a rhythm component in the synthesis scheme?

The papers by Öhman et al. and by Allen concern themselves with production models in general. Öhman's et al. paper argues for a gesture theory of speech production. The authors claim that "the linguistically functional, intended acoustic effects are not, in general, required to have any particular duration; ...acoustic segments with quasi-stationary qualities will arise not as a final end of the phonetic action but as a secondary consequence of the effort to reach a certain final end (the simultaneous sounding of the effects in question)". Öhman and co-authors maintain that the phonological contrast between Swedish words like vila and villa can be eliminated using this analysis. Namely, the stress effect, which takes relatively long to produce, is coarticulated with the vowel /i/ in vila--thus making the quickly producible /i/ long, while the stress is coarticulated with the sequence /i . l/ in villa, thus making the /l/ long.

I would like to ask the authors--if they were here--how they would handle contrasts between long and short vowels in unstressed position--contrasts which are found in a large number of languages, e.g. in Czech and Hungarian.

Allen's paper draws a useful distinction between descriptive models and theoretical models of speech timing, and makes the intriguing prediction that theoretical models may be about to undergo substantial modification, primarily due to the emergence of an "action theory" of speech production. According to that theory, neural activity is hierarchically organized into successively higher levels of coordination, until the highest level of all can only be described in terms of the overall goal of the action. The models of "intrinsic timing" which Allen describes seem to operate at levels higher than a segment; I would like to ask Allen, too, how the segmental short-long opposition can be handled within these theories. It would have been quite interesting to hear some discussion about the almost diametrically opposed approaches taken

in the papers by Allen and Öhman et al. Öhman, as you may recall, states that manifested segmental durations are generally secondary consequences of the effort to produce simultaneous acoustic effects. Thus there appears to be no room for temporal programming as such. The models Allen refers to claim that intrinsic timing is an inherent property of the speech act. Can these two views be reconciled, or will one of them be proved wrong?

Among the papers devoted primarily to perception, Nootboom's presents a decision strategy for the disambiguation of vowel length in Dutch. The strategy presupposes knowledge on the part of the listeners of temporal regularities of speech, and the ability to shift an internal criterion--the boundary between long and short vowels--depending on the speech context. For example, the listener is assumed to know that vowels followed by pause are generally longer than vowels followed by a consonant; that vowels are longer when that consonant is a fricative than when the consonant is a plosive; that vowels are shorter with increasing number of unstressed syllables following the syllable containing the stressed vowels, etc. Nootboom hypothesizes that listeners do indeed possess this knowledge and shift the perceptual boundary between long and short vowels according to speech context. The data presented by Nootboom are quite impressive; it seems to me, however, that there is something artificial in the described situation. When the listeners adjust the criterion depending on the speech context, they are in fact perceiving the total speech act, not just the vowels. Otherwise there would be no need to perform the adjustment. The environment is just as much part of the percept as the vowel. From my experience with English, I would predict that the durations of vowels and postvocalic consonants stand in a compensatory relationship, and that both are related to the overall duration of the word. Even though the strategy Nootboom proposes is quite complex, I submit that it is actually an oversimplification.

Fujisaki and Higuchi present an analysis of the temporal organization of segmental features in Japanese disyllables consisting only of vowels, and find that although the onsets of the transition for the second vowel are distributed over a relatively wide range, a perceptual analysis of the onset of the second vowel shows relatively little temporal variation. It thus seems that the apparent diversity of the onset of transition in various disyllables

is introduced for the purpose of maintaining the uniformity of perceived duration of segments. Fujisaki and Higuchi consider their results supportive of a model in which the motor commands and the articulatory/acoustic realizations of successive segments are programmed in such a way that the perceptual onsets of successive segments are isochronous.

I am quite impressed and convinced by these results and would really like to have more information. Japanese and English appear to have quite different temporal structures at the sentence level. How far does isochrony go in Japanese? Is the disyllabic sequence conceivably a basic unit of temporal programming--for example, if we have a word of four syllables, does it have the length of two disyllabic sequences? Is there any interaction between segments and syllables--for example, how would the inclusion of consonants in the disyllabic sequences influence their duration both in production and perception?

The paper by Huggins is mainly concerned with the intelligibility of temporally distorted speech. Huggins finds that a distorted timing pattern (which often characterizes the speech of the deaf) is a sufficient cause for catastrophic loss of intelligibility. While I have no argument with this particular claim, I would like to take issue with a statement concerning the relationship between pauses and other cues employed to indicate syntactic boundaries. Huggins states that boundaries that are marked by pauses need not be inferred from more subtle cues. In some recent work of mine on the perception of sentence boundaries, I found that listeners can completely ignore a fairly lengthy pause, if it is not preceded by a certain amount of preboundary lengthening and/or change in fundamental frequency. I wonder if Huggins would really persist in claiming that pause is a sufficient boundary signal?

The paper by Donovan and Darwin deals with the perceived rhythm of speech, with special consideration of the problem of isochrony. Their paper tests, among others, a hypothesis that I had formulated in 1973 and discussed in more detail in 1977. My observation was that listeners tend to hear utterances as more isochronous than they really are, and that listeners perform better in perceiving actual durational differences in non-speech as compared to speech. I concluded from this that isochrony is largely a perceptual phenomenon. Donovan and Darwin have confirmed

these results. They make two points in addition: first, that isochrony is a perceptual phenomenon which is not independent of intonation, and second, that it is a perceptual phenomenon confined to language, reflecting underlying processes in speech production. Donovan and Darwin question the value of seeking direct links between syntax and segmental durations rather than indirect ones by way of an overall rhythmic structure.

While I am in enthusiastic agreement with this particular conclusion, I would like to question the presumed role of intonation in establishing the rhythm of spoken language. There is recent evidence (De Rooij 1979) that intonation contributes very little, if at all, to the temporal structure of a sentence: perception of the temporal structure is not noticeably changed when the fundamental frequency is changed to a monotone. In some unpublished work I found that syntactically ambiguous sentences could not be disambiguated by manipulation of the fundamental frequency, whereas they could be successfully disambiguated by systematic changes in the time dimension. (This latter result has appeared in print: Lehiste, Olive and Streeter, 1976.) If Donovan and Darwin persist in their claim, I would like to hear stronger arguments than have been presented in their paper.

The discussion will be structured as follows. The authors will now have approximately five minutes each to make corrections and additions to their papers. Then we will have a panel discussion, lasting about 30 minutes, during which I hope the authors will respond to some of the questions I have brought up--as well as contribute questions of their own that we will all discuss. The last hour of the session will be devoted to a general discussion with participation from the floor. If there is time, I shall try to verbalize some of the final conclusions that emerge from the discussion.

#### References

- Jakobson, R., C.G.M. Fant, and M. Halle (1952): Preliminaries to speech analysis, Cambridge, Mass.: MIT Press (tenth printing 1972).
- Lehiste, I. (1973): "Rhythmic units and syntactic units in production and perception", JASA 54, 1228-1234.
- Lehiste, I. (1977): "Isochrony reconsidered", JPh 5, 253-263.
- Lehiste, I., J.P. Olive, and L.A. Streeter (1976): "Role of duration in disambiguating syntactically ambiguous sentences", JASA 60, 1199-1202.

De Rooij, J.J. (1979): Speech punctuation. An acoustic and perceptual study of some aspects of speech prosody in Dutch, Dissertation, Utrecht.

#### COMMENTS FROM THE PANELISTS

[Since it is impossible to reproduce here the slides shown by several of the discussants, those parts of their presentations that refer to slides have been edited to make them reasonably comprehensible without visual aids.]

R. Bannert reiterated his conviction that the domain of quantity patterns in Standard Swedish and in a number of other languages is the stressed vowel and the following consonant, and questioned the claim that the syllable boundary falls in the middle of a long consonant. He also presented additional evidence concerning the effect of sentence accent on the durational structure of words like stöka and stöcka. Sentence accent lengthens not only the durations of the segments which make up the sequences, but it lengthens all segments of the word in focus, including the second, unstressed vowel of the test word. The segments /s/, /t/ and /a/ have the same duration in both types of test word. The clear difference between the two minimally contrastive words is in the VC sequences of complementary length. The significance of the VC sequences has also been confirmed by perceptual experiments.

D. H. Klatt formulated some general questions that relate to the problem discussed in his paper: 1) what are the phenomena to be described in a particular language, 2) how do all the rules interact, 3) what is an appropriate underlying representation for an utterance in a particular language, if one wants to predict durations or do a complete synthesis by rule? In a linguistics framework, one would like to start with an as abstract--but psychologically real--representation as possible. As regards the rhythm component, it is true that the paper makes the impression that no attempt has been made to account for it; but there are some rules that make the segmental patterns tend to be isochronous, such as cluster shortening rules and polysyllabic shortening rules within words (but not within feet). These two rules, and perhaps some interactions of other rules, bring about a tendency toward isochrony.

B. Granström pointed out that the primary aim of their paper was not to evaluate Klatt's rule system, but to look into what things are important in rule systems in general, and how naturalness of a rhythmic structure is related to intelligibility. Isochrony in perception is obviously there, or the observation would not have been made in the first place; the question is how important it is in production. It might be that it is not even desirable to have isochrony in production. Parallel studies of rhythm in music indicate that music generated by computer with perfect isochrony is often very dull. Another reason why we believe isochrony is not necessary in the description of durational structure is that it turned out that the rule system is actually very good: in the evaluation process, the utterances generated by the rule system were evaluated as being more natural than the actual productions by Dennis Klatt! And measurements show that the output of the rule system was more isochronous than the actual productions. We believe therefore that an isochrony component is not needed, at least not for the generation of the types of isolated sentences produced in our experiment.

G. D. Allen asked how one should handle short and long quantity in intrinsic timing models. According to the extrinsic view, the motor plan includes temporal features which are used by an extrinsic controller (a "speech clock"), which somehow signals the motor system when to begin and end a specified activity. In the intrinsic view, however, the temporal properties of the act are never specified as such but rather are the result of other, not specifically temporal properties of the act. As an example, consider long versus short vowels. An extrinsic timing model would deliver the command to produce the segment (e.g. /a/) along with a "start" command and a durational feature, which would be used by the clock to generate a "stop" command. An intrinsic timing model, on the other hand, would select either the short or long /a/, which must be represented as distinct acts within the motor repertoire, and that short or long /a/ would then be produced as an integrated part of the overall syllable, word, and/or phrase. Its resulting duration would be a complex function of the several interacting levels of structure and behavior which all together define the act.

Asking how one might test for the existence of intrinsic versus extrinsic timing, Allen reviewed an experiment by Laver (cf. J. Laver's comment below) as an example of a potentially useful experimental paradigm.

S. G. Nooteboom presented some data showing that the perceived boundary between short and long vowels shifts in accordance with speech production regularities. The listener has at his disposal a very detailed knowledge of the temporal regularities of speech: he knows how speech should sound in his language. It is more difficult to know how the listener uses this knowledge, and even more difficult to know how it is stored. In the paper, Nooteboom had made a proposition that all this knowledge is stored as a set of rules in the brain, and that the listener rapidly calculates the expected durations of both short and long vowels, places his criterion in the middle between these two, and thus adjusts his judgment according to context. He considers this now to be a very unlikely procedure, mainly because it must be time-consuming to do so much calculation, and also because he does not believe that all these higher-order effects are going straight back to the level of phoneme decision. There is another way of accounting for the same data, in accordance with some psychological models of word recognition.

H. Fujisaki stated that the motivation for his contribution to this symposium was to provide some quantitative means and frameworks for discussing temporal relations within speech units. The successive units of connected speech manifest themselves not as discrete, separable acoustic events, but rather as overlapping and mutually interfering events. Thus, for example, in discussing the issue of isochrony, one cannot claim that a certain point represents the timing of a speech unit just by looking at the speech signal waveform or its spectrogram. In order to decide whether isochrony is a characteristic of speech production or of speech perception, experimental techniques are needed that allow one to infer the timing of the production of segments as well as the timing of their perception. In his paper, Fujisaki showed quantitative techniques to determine these timing relationships. Thus his contribution was concerned not only with perception, but also with production. The material was deliberately restricted to disyllabic two-mora words of Japanese, since they can be regarded as the smallest examples of connected speech. The materials were further

restricted to disyllabic words consisting only of vowels (which are quite common in Japanese), since the articulatory transition from a vowel to the following vowel can be most clearly observed and analyzed from the trajectory of formant frequencies.

Presenting several slides to illustrate the points made in the paper, Fujisaki pointed out a rather wide range of distribution of the onset of articulatory transition among utterances with different combination and order of vowels. At the same time, a strong negative correlation was found between the onset time of such a transition and the rate of transition. In other words, slower transitions were almost always initiated earlier, while faster transitions were almost always initiated later. The onset was distributed over the range from 90 msec to 150 msec within a total utterance duration of approximately 300 msec, which is at least several times larger than the DL for the perception of temporal differences at these durations.

The determination of perceptual timing is based on listening experiments using the same speech material, but by truncating the waveform at various points and presenting only the initial portions as stimuli. The time instant corresponding to 50% judgments was defined as "the perceptual onset" of the second vowel (syllable). The perception of the second vowel starts not at the onset of the formant transition, but at some point where more than 60-70% of the total formant transition has been traversed. The perceptual onsets of the second vowel in various disyllables are concentrated within a very narrow range (about  $\pm$  one DL) centered around the midpoint of the utterance. Thus the initial and the final vowels are almost always perceived as being of equal duration within a vowel disyllable. The results indicate that the isochrony in this case is neither a mere illusion nor a perceptual distortion of the acoustic reality, but the timing of perception actually occurs isochronously. These findings may be interpreted in the light of a model for the control of speech timing (cf. Figure 7, p. 281 of Volume II). One may safely assume that the articulatory control under ordinary utterance conditions is open-loop control. The findings of this research support the hypothesis that motor commands are programmed in such a way that the perceptual durations of the two vowels within a disyllable are perceived as equal, at least as far as Japanese vowel disyllables are concerned.

In reply to Lehiste's questions, Fujisaki remarked that the work is presently being extended into two directions. One is the case of sequences of three or more vowels which are also quite common in Japanese. Preliminary results indicate that the same conclusion holds for these polysyllabic words. The other direction for future study is to include CV-syllables. It is necessary, however, either to establish an analysis technique whereby one can infer from the speech signal the exact timing of consonantal articulation, not just its acoustic consequences, or to rely on physiological observation to determine the timing of speech production and compare it with the timing of speech perception.

C. J. Darwin recalled the purpose of the reported experiment: to distinguish perceptually between two models for the production of speech durations. According to one model, each phoneme has a sort of "platonic" duration which is shortened as a function of syntactic influences; according to the other, there is an underlying rhythmic structure which is perturbed on the basis of the incompressibility of the elements that one is trying to fit into it. The prediction from this theory is that we are aware of the underlying regular rhythmic foot rather than its surface manifestation.

Darwin also presented additional data which supported the claims made in the paper--that people perceive rhythm to be more isochronous than it really is, and also that this does not apply to non-speech. Additional work has been done at Sussex addressing the question whether syntactic boundaries are signalled just by phrase-final lengthening or by lengthening the whole foot in which the boundary occurs. The results show that the latter is the case.

#### DISCUSSION

I. Lehiste recalled the results of some of her earlier experiments which had shown that speakers can use several strategies to signal syntactic boundaries. The strategies have a common result, namely lengthening the foot containing the boundary. These experiments had not tested the relative importance of the different strategies, e.g. of phrase-final lengthening, as boundary cues. Lehiste challenged Klatt and Granström to respond to Darwin. In the discussion which followed, it emerged that even though lengthening of the foot is of primary importance, it does matter what part of the interstress level is lengthened: listeners feel

uncomfortable if the lengthening is limited to the part that follows the syntactic boundary. It appears that both phrase-final lengthening and lengthening of the foot are necessary for listeners to identify the position of a syntactic boundary.

G. D. Allen commented that it is perhaps wrong to call isochrony in English "largely perceptual" (as had been done by Lehiste), since speech is already temporally highly structured in production. He also questioned those of Darwin's results that showed that non-speech was not perceived as more isochronous than the stimuli really were. This finding appears to be at variance with previous research on time perception, and Allen therefore asked (1) was there in fact a trend in the right direction which was smaller than the one for speech and not statistically significant, and (2) what would be the effect on the nonspeech temporal interval perceptions of filling the intervals with various sounds, as the intervals of speech are filled?

C. J. Darwin responded saying that one of the nonspeech results did depart significantly from actual durations, but it went in the other direction--it was perceived as significantly less isochronous. Darwin agreed with the need to perform experiments with different kinds of nonspeech controls with filled intervals. He would also like to perform similar experiments with music.

I. Lehiste expressed the hope that temporal patterning in other languages besides English and the Scandinavian languages might be considered during the discussion, and urged the discussants to remain conscious of the general theme of the symposium: what are the units within which temporal structures are manifested, how does sentence rhythm relate to the durations of these smaller units, and how does sentence rhythm relate to nonphonological aspects of language--e.g. to syntax.

B. Granström found that perhaps too much attention had been given to isochrony in the discussion, and presented some data that showed that a word can be a very important unit for temporal programming.

P. L. Divenyi, referring to his 1977 dissertation, stated that he had found context effects in rhythmic perception in music. If there is no isochrony in the microscopic sense, there could be in the macroscopic sense, even for nonspeech. Rate is a variable that can affect rhythmic perception. Isochrony is an inherent



property of the production system; one could relate isochrony found in perception to production by simply postulating certain listening habits. Thus he does not see any contradiction between productive isochrony and perceptual patterns found in perceptual experiments.

L. Lisker suggested an experiment: to assign segment durations by a random process (in synthesis), and find out what loss in intelligibility and naturalness there would be.

R. Gsell discussed temporal relations in Thai, a quantity and tone language. Stress has a leveling effect on quantity contrasts. Temporal constraints and perceptual limitations produce for the listener neutralization of contour tones in shortened and unstressed syllables.

E. Selkirk took issue with the moderator's characterization of generative phonology as a theory which is in principle unable to countenance such notions as syllable, timing, and rhythm. The notion of the phonological representation within the theory was one of a purely linear kind which saw it as a sequence of segments and boundary elements. In recent years, though, workers who see themselves as operating within the context of generative phonology have been rediscovering that this conception of phonological representation has to be radically revised, allowing for far richer hierarchically arranged suprasegmental structures.

Some workers, Selkirk included, have been arguing for a rather different conception than that in the Sound Pattern of English of Chomsky and Halle, of the relation between phonology and syntax in a generative grammar. In this conception, syntax is seen as bearing on phonology only insofar as phonological units, like syllables or intonational phrases, may have specific syntactic domains over which they are defined, but phonological and phonetic processes are seen as functioning only in terms of these phonological hierarchical structures. It is a claim of this theory that something like final lengthening has its domain defined in terms of phonological units (such as intonational phrase and perhaps others); it would not be immediately sensitive to syntactic structure. What is predicted here is that there would be a systematic convergence of various types of phonological phenomena; the unit at the end of which one finds lengthening would be the same one with which, for example, an intonation contour would be associated, or it may also be the domain of rules of segmental phonology.

Lengthening or the realization of intonational contours and so on are not conceived as individually and separately sensitive to units of syntactic structure.

H. Fujisaki, responding to comments by P. Divenyi and L. Lisker, agreed that we need to look at both microscopic and macroscopic levels of timing. There should be a hierarchy of levels in which speech timing is programmed and maintained. For instance, the problem of compensation between the duration of a consonant and the following vowel is a matter of timing within a syllable, but the compensation between the duration of a vowel in a CV syllable and the following consonant of the next syllable is a matter of interaction between sub-syllabic units across syllable boundaries. Fujisaki had looked at vowel disyllables in order to investigate the relationship between durations of the two syllables without having to consider the problem of consonant-vowel compensation.

J. Laver reviewed his "motoric balance point" experiment mentioned by Allen in connection with two opposing views of the nature of the control of temporal relations. The argument is between the extrinsic view of temporal control, where a "speech clock" acts as an external, overlaid control device, versus the intrinsic view, where temporal relations are the direct product of characteristics of segmental representations themselves. Laver singled out one finding in his experiment which tends to support one of these views. When his subjects were faced with the need to produce forms which had a quantity difference as well as a quality difference between them, such as PEEP and PIP, then the link between quantity and quality was very labile in their productions, and very easily perturbed. There were many errors made, where the right quality but the wrong quantity was produced. So there were examples of PIP with a long vowel duration and of PEEP with a short vowel duration, where both nevertheless showed appropriate articulatory quality. This tends to support the extrinsic view, where duration is at least to some extent the product of specific neuromuscular programming separate from programming for articulatory spatial targets as such.

N. Thorsen addressed a question to Nooteboom, who, with his last slide, had appealed to the audience to have the courage to assume that word identification precedes phoneme recognition. Thorsen asked how Nooteboom would account for the perception of

slips of the tongue, which are generally perceived as such, i.e., as slips or mistakes, while at the same time the word is being identified correctly.

K. L. Pike, in his comments, made the point that in English, both isochronic and non-isochronic timing are essential. Under certain circumstances, we must not have isochronic stress groups; under other instances we must indeed have them. This is connected with the fact that in his normal use of English there are some items which one might call "double stresses". These are, in general, related to certain kinds of syntactic groups. There is also a kind of a semantic component which often goes with these double stresses. It is a unitizing effect, tying the items together in some kind of a single concept to be viewed as a unit rather than as components loosely strung together. We must not be so inflexible that we assume that we must have either isochronic stress groups or else we must have largely non-isochronic stress groups. In Pike's analysis of the material one must leave room for both in English. This, in its turn, forces another conclusion: we cannot assume that there is a single rigid set of rules mapping directly, and in only one manner, material from the grammatical hierarchy on to the phonological one; nor of semantically oriented units from a referential hierarchy on to the grammatical or phonological one. We need three hierarchies, always interacting one with another, but never the one totally determining the other. Our rule systems, therefore, cannot be inflexibly from grammar to semantics and phonology; nor from semantics to grammar and then phonology. Rather we must have some interdependence in which the purpose of the speaker is distributed in ways which are vastly more complex than a one-way rule system can tell us.

S. Nooteboom, responding to Thorsen, disclaimed having ever implied that listeners cannot extract phonemes from the acoustic signal. In the normal recognition of known errorless words--which is usually very fast indeed--it is not necessary to assume that phonemes are mediating in perception. Hearing unknown words, or words containing detected mispronunciations, listeners must have been listening in a "phoneme mode".

L. Nakatani questioned the existence of isochrony in production. Even though in comparing black dog with blackish dog there seems

to be isochrony, this can very easily be explained by the fact that the first syllable in a bisyllabic word becomes shortened relative to the same syllable in monosyllabic words. There is another factor operating here--resyllabification. In blackish, the /k/ is aspirated, indicating that the /k/ now belongs to the second syllable. So one cannot compare black and blackish, for the syllables are different. If one controls for this by using reiterant speech, some kind of compensation can indeed be found; but if one controls for that and looks at the effect due to the insertion of an unstressed syllable in medial position, one does not find any compensation. Similarly, if one inserts an unstressed syllable at the beginning of the second word, there is no compensation. There is a very linear relationship between the number of intervening unstressed syllables and the interval between stressed syllables. This is consistent with data collected by Wayne Lea.

Nakatani has also looked at duration patterns of words in different contexts. If there is a tendency toward isochrony, the durations of words should vary as a function of the context in which they occur. Looking at the same words in different positions in different sentences, Nakatani found that the duration patterns of words were extremely consistent, and concluded that there is no evidence for isochrony in production. Therefore it should be ascribed primarily to perception, and be based on the fact that content words and function words alternate, and that most bisyllabic words in English have the stress on the initial syllable.

I. Lehiste remarked that there are usually several principles operating at the same time, and they interact. Tendency toward isochrony is one of these principles, but there is certainly another one--the principle of maintaining the temporal integrity of the word, so that the duration of a monosyllabic word is roughly comparable to the duration of a disyllabic word. When these two principles interact, they will influence each other.

E. Uldall noted that we are devoting our attention almost entirely to "stress-timed" languages (though there have been references to Japanese). She expressed the wish to hear a lot more about the opposite case: for example, about French. Phoneticians very frequently refer to English as a classic case of stress-timing, and to French as a classic case of syllable-timing. Yet all the experimental evidence we have about English shows that the



"rhythmic feet" are far from isochronous, and what Uldall has seen of French syllables makes her think that they are not isochronous either. So why do phoneticians go on saying what they do?

G. Fant stated that most of our data about durations have been obtained from speech waves--oscillograms and spectrograms. The question is, can we interpret this in terms of a production model to give a better perspective? The answer is affirmative. For instance, if we study vowels in sentence-final stressed position, we find that all the durations are the same, because what has determined the termination of the vowel is the phonatory gesture which is the same for all vowels and independent of the preceding consonant. On the other hand, if the vowel is followed by a consonant, the consonantal frame influences the vowel duration. This is the articulatory aspect. So the duration of a vowel can be set either by phonation or by articulation or, really, both. If a voiced stop comes after the vowel, then of course the vowel is terminated as the acoustical consequence of the constriction, but if it is an unvoiced plosive which comes after the vowel, then there is a separate neural command for the abduction of the vocal cords. That command is somewhat time-locked to articulation, but they are still separate events. This can be a fruitful way of scrutinizing the durational data.

S. M. Marcus gave a brief summary of his research concerning Perceptual Centres or P-centres, which involve rather more fine-grain aspects of speech timing than those determining the temporal structure, isochronous or otherwise, of continuous speech. In producing perceptually isochronous sequences of isolated monosyllables, perceptual regularity corresponded to no simple physical alignment. Subsequent experiments have shown the P-centre locations to be a function of the acoustic structure of the whole stimulus--for example extending the /t/ closure of "eight" shifts its P-centre. These results clearly demonstrate that before considering such questions as isochrony and "syllable-" or "stress-timing" in continuous speech, we need to be very clear what we are measuring the timing of. We must be wary of assuming that simple instrumental measurements, such as consonant and vowel onsets and durations, are related in other than a complex way to our perception. We should also be aware that much of the data which has been

used to demonstrate either isochrony or lack of isochrony now needs to be carefully reexamined.

G. D. Allen urged the audience to view timing and rhythm as mental phenomena. Time as it is measured in spectrograms and oscillograms is but one correlate of timing and rhythm. These phenomena belong in the mind, several levels removed from the articulatory periphery.

I. Lehiste thanked the panelists, the very efficient chairman, and all contributors from the floor. She observed that many issues had remained unsolved--for example, the question whether isochrony in English is a property of production or perception. One underlying assumption, however, appears to have been generally accepted--namely that temporal organization operates within units that are larger than a single segment. The task still remains to establish these units for different languages. She concluded with the hope that this discussion has contributed some background that will be taken into account in future research directed toward the discovery of the temporal structure of language.

## SYMPOSIUM NO. 6: MOTOR CONTROL OF SPEECH GESTURES

(see vol. II, p. 315-371)

Moderator: James Lubker

Panelists: R.A.W. Bladon, R.G. Daniloff, Hajime Hirose, Peter F. MacNeilage, and Joseph Perkell

Chairperson: Leigh Lisker

## JAMES LUBKER'S INTRODUCTION

In preparing my introductory comments for this symposium I have made two assumptions: first, I am assuming that those of you in attendance are interested in speech production/motor control theory and have therefore taken the time to at least glance through the papers for this symposium as they were published in volume II; and secondly, I am assuming the goals of phonetics to be as described by Björn Lindblom in his plenary lecture (p. 3-18, this volume).

Acceptance of the first of these assumptions implies that I need not spend much time in summary of the papers in this symposium; they are there for the reading. Rather, I will take as my goal to provide a common framework for those papers and the points of view expressed in them, in order to allow the discussion of current and important issues in production/motor control theory.

Since acceptance of the second assumption will dictate the nature of the framework and issues which we will develop for discussion, it is perhaps wise for me to be somewhat more explicit about it. In the summary (vol. I, p. 3-4) Lindblom states: "Phoneticians accordingly construe their task of speech sound specification as a physiologically and psychologically realistic modeling of the entire chain of speech behavior." And he then goes on to pose the questions of (1) why it should not be possible for "phoneticians to extend their inquiry into the sounds of human speech to ever deeper physiological and psychological levels using speech as a window to the brain and mind of the learner, talker and listener?", and (2) "Why we should not expect more complete, theoretical models and computer simulations to be proposed for speech production, speech understanding and speech development that match the present quantitative theory of speech acoustics in rigor and explanatory adequacy?".

Indeed, the very title of this symposium, The Motor Control

of Speech Gestures, suggests research and theory devoted to an attempt to elucidate the rules and systems "at ever deeper physiological and psychological levels", by which man generates speech, and to do so with as much precision and scientific rigor as possible. Motor control research and theory must be integral to the goals stated by Lindblom, that is, to the development of explanans principles in phonetic and linguistic theory. Thus, the acceptance of those goals is my second assumption for this symposium.

There remains, however, much room for discussion since the search for precise and valid explanans principles for the generation of human speech is currently faced with several crucial issues, which are well illustrated by the papers presented in this symposium. Those issues can be discussed within three very broad and highly interrelated areas of theory and research.

In the first place, many questions in motor control/production research have quite naturally dealt with the form and function of the system or systems which operate to produce a speech acoustic signal. That is, a major effort in motor control research has been the attempt to discover the rules which explain and predict the transformations at the several interfaces in the chain of language generation and perception. Armed with such rules we would indeed have "a window to the brain". And since that is precisely where language resides, knowledge of these rule systems would provide us with a strong tool for the elucidation of certain aspects of language theory. Efforts to discover the rules have not, thus far at least, resulted in a Motor Control version of the Acoustic Theory of Speech Production, but as Lindblom suggests, there is no reason to believe that we will not one day have such a theory. Every paper in this symposium deals via proposed models, specific data or both with the form and content of such rule systems and it would thus seem obvious that this should be a fruitful area for discussion.

A second broad area of theory and research in the motor control of speech gestures is the precise form or nature of the units which serve as input to the motor control systems. In the papers of this symposium a number of possibilities are suggested: Abbs uses a matrix of phonetic features; in an updated version of their paper Daniloff and Tatham also suggest such a matrix. Bladon

considers several possibilities including features, phonemes and phonological syllables; Gay and Turvey seem to be viewing the input as phonemic; Perkell agrees that studies of motor control mechanisms are closely related to the nature of the "fundamental units underlying the programming of speech production", but he does not speculate in this paper as to what those units might be. Although the papers of Folkins, Hirose, and Sussman are concerned with specific experimentation with the functioning of the motor control systems, irrespective of the input unit, the nature of that unit would clearly seem to be a second broad area for useful discussion.

Finally, let me propose a third general area for discussion; an area which is so related and intertwined with the preceding two as to be virtually inseparable from them. It concerns more the form of attack upon the problems of the preceding two areas.

I have been implying that motor control rules of some kind are necessary in order to move from abstract linguistic concepts such as the phoneme or syllable to the concrete data obtained in speech production experimentation. These two sets of units, the abstract concepts of linguistics and the hard data of production research have never been very well matched and if they are to be used together in attempts to explain speech and language generation then transformation rules would, in fact, seem necessary. Fowler et al (1978) have called such efforts "Translation Theories" and they contend that virtually all production research to date may be classed as one or another type of translation theory. Fowler et al also suggest that all abstract linguistic units possess three properties: they are discrete, static, and context-free; while all units of production are dynamic, continuous and context-adjusted. A clear mis-match! Most of us would agree with Liberman and Studdert-Kennedy (1978) that translation from discrete, static and context-free to dynamic, continuous and context-adjusted requires a "drastic restructuring" of segments, whatever the original input segments might be. Thus, the many attempts to provide theories which explain and solve the non-isomorphism between the abstract linguistic units and the concrete production units. In the course of that work much effort has been expended toward attempts to find physical/physiological correlates of the abstract linguistic units... to eliminate the non-isomorphism.

To date this research has been notorious for its lack of success and physical/physiological correlates of abstract linguistic units are conspicuous largely via their absence. Such repeated failures have caused some researchers to become disenchanted with the particular research strategy entailed in translation theories. They contend that when experimental data are shown repeatedly to be at variance with theoretical constructs it is only natural to begin to question the legality of the constructs. Carried on, such an argument raises the question: should production/motor control theorists develop their own units and concepts which are based on actual experimental observations of motor control mechanisms in general and which are unbiased by notions and abstract concepts borrowed from linguistic theory? Moll, Zimmerman and Smith (1977) have presented perhaps the most explicit and extreme version of this view and they suggest that: "Such an approach might lead us to the identification of units of programming based on the physiological parameters of movement, muscle contractions and neural activity, units which might or might not correspond to any construct previously defined."

Although such a view may be compelling, it can lead to a small feeling of scientific schizophrenia in those of us who have for so long followed the "translation theory road". The notion of sets of transformation rules between such interfaces as the output of a phonological component and the neurophysiological structures of the speech producing mechanism seems such a reasonable notion. The linguistic concept of "phoneme", for example, is indeed an abstract one... unseen and unseeable. But so also are many of the concepts of the physicist unseen and unseeable. Further, Fromkin and others both previously, and here at this Congress, have discussed persuasively the psychological reality of linguistic units as demonstrated by, for example, speech errors. Nevertheless, the arguments proposed for not allowing ourselves to be prejudiced by the use of preconceived and abstract linguistic notions may also be persuasive and there may thus be some benefit in discussion of this issue.

In any case, we see two quite differing points of view concerning the theoretical and experimental approach to the general problem areas of input units and motor control rules and systems. And, there is yet a third point of view. Bernstein's Action Theory

(1967) was originally proposed as a general theory of coordinated movement. Turvey (1977) and his associates (e.g., Turvey et al, 1978; Fowler et al, 1978) have applied this theory to the generation of speech and language. The action theory point of view also argues against the use of translation theories in speech production/motor control research, but does not agree that such research should be conducted without reference to linguistic units. These investigators' use of action theory and their development of such concepts as "coordinate structures" in speech motor control represent an attempt to avoid translation theories while at the same time not rejecting out of hand the use of all traditional linguistic concepts.

And so, the problems regarding our experimental approach to the nature of the input units and the motor control rules and systems which act upon those units would seem to be: (1) Should production/motor control theorists continue to search for translation rules which mediate between abstract linguistic units and concrete production units, or (2) Should production/motor control theorists attempt to ask questions about fine motor behavior in general in an attempt to elucidate speech and language generation and in the process create new or substantiate old input units, or (3) Should production/motor control theorists follow the entirely new course proposed by Action Theory and its claim of understanding linguistic organization via experimental study of the lower, "basic" properties of speech acts without the use of translation rules? I should add, since there was some misunderstanding at the symposium, that I have here only stated these as experimental approaches worthy of discussion and I have not aligned myself with any of them in this paper.

It seems to me that this symposium offers a reasonable forum for the discussion of these very important issues.

Here, then, are three very broad and interrelated areas of research and theory from which we might profitably draw questions for discussion: (1) the nature of the programming units; (2) the form and structure of the system or systems which act upon those units; and (3) what the best theoretical approach might be to discover what those units and systems are.

Each of the papers in this symposium takes up issues in one or more of these broad areas and it may now be appropriate to

consider some of their specific points of view.

For example, one topic which may be of general interest to all of the papers and which may involve each of the three areas discussed above is: What is the nature and the relative roles of feedback mechanisms versus central programming/simulation loops in motor control systems?

In that framework Abbs presents a model which stresses that not only is afferent feedback required in speech control, but it must take place at a variety of sites, including rather low level ones, in order to account for speakers' ability to compensate rapidly to unanticipated disturbances in ongoing speech. While he does not reject out of hand the possibility of a pre-adjustment, or efferent copy, system he argues that afferent control capability is the prime factor in accounting for rapid adjustments to dynamic unanticipated loads.

Perkell, on the other hand, argues that both orosensory feedback and central programming with internal feedback play important roles in motor control. Specifically, he implies a major role for central programming and internal feedback (feedback entirely internal to the central nervous system) "for the moment-to-moment (context-dependent) programming of rapid movement sequences".

Gay and Turvey present still a third possibility in the form of data which they interpret as being negative to the existence of an open-loop control system and positive to the function of the coordinate structures of Action Theory. Their principle argument against any closed loop system, "internal" or otherwise, is that "while an error signal can index how near the collective action of a number of muscles is to the desired consequence, it does not prescribe in any straightforward way how the individual muscles are said to be adjusted to give a closer approximation to the referent."

Several of the papers present data which are relevant to these theoretical observations. For example, in one experiment Folkins provides an indication of the variability, and thus the trade-off in muscle function, for jaw elevation, thereby supporting MacNeilage's (1970) earlier views on the variability of muscle activity for the attainment of particular vocal tract targets. Additionally Folkins shows that the medial pterygoid muscle contracts in a similar manner with or without a bite block in place thus

suggesting that "unnecessary" jaw closing activity is not eliminated either in the equations of constraint proposed by Action Theory or in the central movement plan of a simulation loop.

Data supportive of intermediate stages of feedback control as well as different patterns of control, which tends to support the model proposed by Abbs are presented by Hirose in his study of electromyographic activity and movement of the soft palate.

Sussman's elegant single-motor unit work demonstrates evidence for cellular level reorganization of muscle function in jaw elevation in response to a "behavioral and biomechanical aspect of the encoding program for speech.

These and additional experimental data provided by Folkins, by Hirose and by Sussman must be considered in the theoretical interpretations provided by Abbs, by Gay and Turvey and by Perkell. Perhaps in doing so, and in discussing additional data, we can make some progress in the question of the nature and relative roles of feedback and central programming. Unfortunately it must be noted, in retrospect, that such a discussion was difficult for the panel to initiate, largely due to the fact that several of the authors were unable to attend the congress. Specifically, Abbs, Folkins, Gay, Turvey and Sussman were not present on the panel. Sussman was ably represented by Peter MacNeilage but it was not possible to get the viewpoints of the others in the form of direct discussion.

Nevertheless, with all of these issues, ranging from the relative merits of translation theory versus action theory versus (for want of a better term) exclusively neurophysiologically based theory to the issues of the relative importance of feedback versus central programming, I think that without any more preambing on my part we have more than enough conflict with which to begin a discussion of the motor control of speech gestures.

#### References

- Bernstein, N. (1967): The coordination and regulation of movements, London: Pergamon Press.
- Fowler, C.A., P. Rubin, R.E. Remez, and M.T. Turvey (1978): "Implications for speech production of a general theory of action", in Language production, B. Butterworth (ed.), New York: Academic Press.
- Liberman, A.M. and M. Studdert-Kennedy (1978): "Phonetic perception", in Handbook of sensory physiology, vol. III, 'Perception', R. Held, H. Leibowitz, and H.-L. Teuber (eds.), Heidelberg: Springer-Verlag.

- MacNeilage, P.F. (1970): "Motor control of serial ordering of speech, Psych.Rev., 77, 182-196.
- Moll, K.L., G.N. Zimmerman, and A. Smith (1976): "The study of speech production as a human neuromotor system", in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.), 107-127, Tokyo: University of Tokyo Press.
- Turvey, M.T. (1977): "Preliminaries to a theory of action with reference to vision", in Perceiving, acting and knowing: Toward an ecological psychology, R. Shaw and J. Bransford (eds.), Hillsdale, New Jersey: Erlbaum Press.
- Turvey, M.T., R. Shaw, and W. Mace (1978): "Issues in the theory of action: degrees of freedom, coordinative structures and coalitions", in Attention and performance, VII, M. Requin (ed.), Hillsdale, New Jersey: Erlbaum Press.

## COMMENTS FROM THE PANELISTS

Two panelists had comments to make on the nature of the programming units. MacNeilage pointed out the potential of single motor unit research as a means for defining the nature of such units, although he also made clear that at present he and his colleagues are not attempting to posit "any straightforward relationship between these data and such concepts as the phoneme or distinctive feature". Bladon spoke somewhat more extensively on this issue. Specifically, Bladon called for the recognition of "a plurality of articulatorily relevant units", including features, phonemes and phonological syllables. He provided examples in support of each of these and then went on to say, "moreover coarticulation needs to be sensitive at times to other properties than phonologists have proposed, including a strength hierarchy, including even rule-order in rapid speech forms, and including also phonetic system size (perhaps implying some sort of articulatory distance measure)". He then noted that the existence of counter-examples against all of these units might "lead into the question of perhaps whether an interesting possibility would be that different types of units might be made use of for different motor control functions".

Two panelists also took up the question of the form and function of motor control rule systems. Hirose directed his comments to these systems by pointing out that his overall aim was to "investigate the temporal organization of the speech production process", via investigations of the "relationship between the pattern of motor control signals...and the dynamic characteristics of the speech organs which act in response to the control signals". In summarizing the EMG and movement data from velopharyngeal func-

tion in Japanese presented in his paper (vol. II, p. 351-357) Hirose noted that both the EMG activity and the resultant velar movement for nasals varies predictably depending upon the class of nasal sound being produced. He states: "It can be assumed that the EMG activity for moraic /N/ is characterized by a step-like suppression and the velar movement can be regarded as a smoothed response of the second order system to it. For the initial /m/, the velar movement can be taken as a ballistic impulse response like movement. For the geminate /Nm/ there must be a positive control which can inhibit extreme lowering of the velum in spite of the longer duration of nasalization." Thus, Hirose stressed the importance of studying the relationship between EMG activity and structural movement as one method for evaluating potential motor control rules and systems. Daniloff and Tatham, on the other hand, investigated EMG activity in the production of English bilabial stops. In a reinterpretation of the original data, Daniloff reached the following conclusions, among others: First, there is "definitely an impression from the data of multiple articulatory solutions (there is no one muscle nor any one articulator that needs to move in exactly the same way from trial to trial to get a given acoustic end) and, thus, you need to know the biomechanics of an articulator in order to interpret the EMG". Secondly, and related to the first point, "coarticulation, which you expect to be extreme in a stop consonant-vowel syllable, may be optional or there may be ways to solve the coarticulation using different muscles from repetition to repetition". Finally, Daniloff stressed the close relationships which they noted between temporal characteristics of their EMG data and the resultant labial productions. Thus, in agreement with Hirose, Daniloff provided examples of the use of relationships between EMG activity and output behavior of the structures.

Two of the panelists presented views concerning the best theoretical approach to motor control research. MacNeilage stated that one of the reasons underlying his interest in single motor unit work "derived from a relative disenchantment with attempts to define the underlying abstract units of the speech production process on the basis of experimental studies of speech production". He thus wanted to provide some data about the rather high level stage of the motor unit, which he believes "defines the way the



central nervous system must encode its information", before ultimately returning to the "larger questions" of underlying units. Bladon, on the other hand, expressed concern that "the limited predictive capacity of each of these linguistic constructs (features, phonemes, and phonological syllables) have led various people to be critical". Specifically, Bladon cited both MacNeilage and Lubker in statements relevant to the lack of correspondence between production data and theoretical linguistic constructs. He suggested that "large numbers of linguistic constructs have been shown to have some relevance to the control of coarticulation and if they have come to very little effect in their operation, can you really expect all data to be supportive of any one construct?" Bladon answered his own question in the negative and expressed considerable unease at the "nihilistic" views of Moll, Zimmerman and Smith (1977) cited above in the introductory comments. In the subsequent panel discussion, MacNeilage extended his views somewhat by stating: "I think the basic state of affairs is that we have a linguistic message that we are trying to implement by a motor control system and the implementation of that message must obviously be related to the nature of that message and therefore we need to continue to struggle with the problem of what the underlying abstract forms are." And further, speaking directly to the issues raised by Bladon, he stated: "When I say that I think the theory is relatively unsuccessful, what I mean is that there is no simple set of rules that can account for the observed coarticulatory behavior. I think our problem is that we just simply have too many divergent pieces of data and we do not have a clear-cut relationship between those data and the underlying concepts like the syllable. So, we have these kinds of anomalies and we have these fairly spectacular cross-language differences in exactly how speakers handle coarticulatory events, and I would stick with my characterization that the theories have been relatively unsuccessful." In return to MacNeilage's comments, Bladon agreed that there was no simple set of rules but did not think "that we should therefore conclude that a complex set of rules is a non-successful one". It would thus seem that both Bladon and MacNeilage were concerned with some form of "translation theory" approach to motor control systems in spite of some differences regarding the nature of the translation theory. Indeed, this seemed to be true in the case of

all of the present panel members. The paper by Gay and Turvey was supportive of Action Theory but since neither of them were present that view was not taken up at this point in the discussion.

Finally, Perkell provided a consideration of the relative roles of feedback and central programming mechanisms in motor control systems and in doing so pointed out that it is necessary that we "understand the way feedback works if we are ever going to come close to understanding the physiological/neurophysiological correlates of linguistic units". Perkell suggested three forms of feedback which might be important to speech motor control: (i) "oral-sensory feedback utilized over relatively long time spans in conjunction with auditory feedback to establish and maintain a subconscious knowledge of certain vocal tract states which produce sound outputs that have distinctive and relatively stable acoustic properties"; (ii) "peripheral feedback used to inform the control mechanism about changes in the frame of reference which must be taken into account in making adjustments in motor programs". Perkell discussed this second point in detail in his paper (vol. II, p. 358-364). In the present discussion he added the notion that "when a motor program is constructed and executed, it is probably accompanied by a set of expectations on the outcome of the program and feedback is likely used to compare the actual with the expected result. If a large enough mismatch is detected then adjustments have to be made in subsequent programs."; (iii) "Feedback could be used on a moment-to-moment basis in the partial control of the individual's articulatory movements or in the coordination of more or less simultaneously occurring movements of different articulators." In discussing this last form of feedback control Perkell brought in the work of Folkins and Abbs (1975) which suggests that the "peripheral reflex pathways are programmed to make on-line or moment-to-moment adjustments in commands to the articulators". He also discussed the work on head-eye coordination in monkeys which has been shown to be controlled by reflex pathways involving the vestibular apparatus. This, in turn, led him to the question: "is there anything like the vestibular apparatus for vocal tract movement coordination? In other words, in what ways might the neural organization for speech production be specialized for moment-to-moment use of peripheral feedback?" Perkell warned that in seeking answers to such questions we must

be very cautious since the experimental conditions in feedback research might cause subjects to use mechanisms which are 'available' but not used for ordinary "ongoing overlearned speech activities". Perkell concluded by suggesting that "a great deal of movement control for ongoing adult speech production is probably accomplished through pre-programming. We use motor patterns which are stored in some kind of incomplete form and elaborated in part during pauses and in part on a moment-to-moment basis. The control mechanism could use what the motor control theorists like to call 'efferent copies' or a knowledge of ongoing motor commands which could be used to compensate for self-generated changes without having to resort to peripheral feedback. In order to account for natural variations in articulatory movement (e.g. motor equivalence) some moment-to-moment feedback function seems to be necessary. Now, this feedback function could include peripheral feedback and it probably includes feedback mechanisms contained entirely within the central nervous system (cf. the discussion by Hirose, below). The use of internal feedback in place of peripheral feedback might be part of learning how to speak and there is most likely a fluctuating use of various forms of feedback depending on the demands of the situation."

In addition to these relatively formal comments there was also some more informal discussion among the panel members, some of which has already been alluded to in the above section on theoretical approaches to questions in motor control. During this discussion Perkell pointed out that coarticulation is observed in terms of structural movement and that "we don't see the movements of features". He further observed that structural movement, using the example of the mandible, is set by goals specified as a function of time and influenced by the movement and positions of other structures such as the lips, tongue body, tongue tip and even the larynx. All of these requirements on the mandibular movement must be summed so that they "produce a set of motor goals for the mandible which is really vertical position as a function of time". Further, what seems to apply "almost universally" for such conditions is some form of "look-ahead" mechanism which checks for future goals and intervening requirements, thus allowing smooth movement from goal to goal. Perkell then notes that recent data (see discussion below by McAllister) suggests that in rounded

vowel-nonlabial consonant-rounded vowel utterances there is a trough, or reduction, in EMG activity that would not be predicted by a look-ahead mechanism. He then called for some discussion of such look-ahead mechanisms and the possibility of word or syllable boundaries to help us "nail down" such data. In response to this, Daniloff suggested that juncture which exceeds some given length of time may result in suppression of activity in certain articulators and movements towards more neutral positions. Bladon noted that although the mass of data seems in favor of articulatory spread of features such as rounding across syllable and word boundaries there may well be cases in which speakers are simply using different strategies and where boundaries "have come to be influential". However, he does feel that the weight of the evidence is to the opposite and that coarticulation does spread across such boundaries.

#### DISCUSSION

Since space does not permit the inclusion of all points made during the open floor discussion, only those points most relevant to the main issues raised by the panel will be taken up. Additionally, priority is given to those who were motivated enough to comply with the Congress Organizers' request to supply written summaries of their questions.

Löfqvist provided an extensive discussion of Action Theory. He pointed out that not much experimental work had yet been done within that framework but that theoretical considerations are equally important and that theoretical arguments and issues should be sorted out before starting experimental work. He said that "one of the main problems in motor control, emphasized by the Russian physiologist Bernstein, is that of reducing the number of degrees of freedom to be directly controlled". He also suggested two problems which any explanatory theory of motor control must deal with: "Movements should be made to reach a given goal irrespective of varying initial position", and "Movements should be carried out in the face of unexpected perturbations or changes in the environment." Löfqvist emphasized that both of these movement conditions must be carried out "without any lengthy search procedure". Action Theory accounts for such movement phenomena via the concept of coordinative structures, which can be "regarded as a functional grouping of muscles constrained to act as a unit."



Specified relationships between a group of muscles, expressed by equations of constraint, make the group self-regulatory." He suggested, in closing, that "the perspective of coordinative structures would lead you to predict that invariance will not be found in the individual muscles. Rather, it should be searched for in the dynamic relationships between muscles, or groups of muscles, over time.

In response to Löfqvist's comments, Lindblom asked how Action Theory accounted for the ability of the motor system to adapt to an almost infinite number of new situations while goals remain constant. Lindblom further called for the panel to clarify the term "pre-programming" which he took to mean, in general, "some kind of adaptive, creative control strategy derived on-line and involving foresight". Specifically, Lindblom called for discussion of a possible mechanism to account for such control. Hirose answered Lindblom's second question by reference to a cerebro-cerebellar loop which has been proposed by Allen and Tsukahara (1974). These authors describe a specific neurophysiologic system, the cerebro-cerebellar communication system, "the function of which is largely anticipatory, based on learning and previous experience and on preliminary, highly digested sensory information that some of the association areas receive."..."In other words, in central monitoring of efference, a copy of the motor commands sent to the muscle is monitored centrally and thus it should not wait for proprioceptive comparison." Bladon also offered some comments on Löfqvist's view of Action Theory and in doing so extended Lindblom's question concerning it. Bladon first stated that he felt that the concept of coordinative structures was quite promising. Nevertheless, he felt that there was a major problem which both Löfqvist and Lindblom had alluded to, and that was, "how do you actually investigate this, how do you test this theory, how do you compare it with what you have already?" Bladon suggested that since it has been stated that coordinative gestures involving speech are agents of coordinative structures, then perhaps experimental proof of the existence of such coordinative gestures would provide the sought after evidence. In reviewing that evidence with which he is familiar Bladon was unable to provide any direct support for such coordinative gestures and feels that the question of experimental proof for Action Theory remains

an unanswered and important one.

Somewhat later in the discussion Port made a comment which was relevant to the Action Theory concept. He argued for a less limited role for timing in coarticulation theory. Specifically, he suggested that "an adequate theory of coarticulatory phenomena should probably also include explanation of examples of inherent durational effects and their compensatory adjustments as an integral part of the system--not as a different theory patched on at the end. It is even possible that by building in this kind of temporal coarticulation at the outset, we will find the entire project more tractable." Port then stated that "the notion of coordinated structure employed in action theory is intended to capture both the temporal and spatial invariants of a phonetic event. Perhaps this is a theoretical notion that could be developed to capture both the temporal aspects of the spatial position of articulators as well as the inherent temporal structure of segments and prosodies."

Turning in another direction, McAllister responded to Perkell's question (see above) concerning the failure of "look-ahead" models to account for the observed "trough" in recently reported EMG data. McAllister showed simultaneous movement and EMG data from labial function during the production of rounded vowel--nonlabial consonant string--rounded vowel utterances. The nonlabial consonant strings consisted of one, four and six consonants. These data clearly showed troughs, or relaxations, in both the EMG activity and in the lip rounding, the most interesting point being that the relaxations occurred at the boundary between the offset of the consonant string and the onset of the second vowel. McAllister agreed with Perkell that such data are incompatible with previous descriptions of the look-ahead mechanism, and stated that he is particularly "hard pressed to explain the location of the trough." He suggested that there may be "a critical acoustic boundary" at that point which demands a "neutralization" of rounding.

Ohala suggested that our search for underlying units would perhaps be facilitated by examining cases where coarticulatory behaviors were "clear" rather than "smeared". Specifically, he presented a number of examples of cases, in Swedish and in English, where coarticulatory behavior was time-locked to phonemes.

As a final point in this summary of the discussion from the floor, the comments made by Porter may be appropriate. Porter called for considering production and perception phenomena more closely together rather than as distinct fields of study. He felt that this would aid us in "terms of understanding perception and also in understanding the role of feedback in the control of output". Porter extended his argument via Action Theory by noting that somewhere between "abstract phonetic entities and the more concrete properties of motion and acoustics" there must be an "interface and a common code". That is, a common code to the exclusion of a translation theory. A code that functions both in production and in perception.

Very little summary is required for the above comments. It seems very clear that answers are being sought and that there is a healthy amount of controversy. The seeking and the controversy suggest that researchers in the field of motor control are, indeed, working toward those goals stated by Lindblom in his plenary lecture: that "phoneticians should extend their inquiry into the sounds of human speech to ever deeper physiological and psychological levels using speech as a window to the brain and mind of the learner, talker and listener", and, further, that we should expect "more complete, theoretical models and computer simulations to be proposed for speech production, speech understanding and speech development that match the present quantitative theory of speech acoustics in rigor and explanatory adequacy".

#### References

- Allen, G.I. and N. Tsukuhara (1974): "Cerebrocerebellar communication system", Physiol.Rev. 54, 956-1006.
- Folkins, J. and J. Abbs (1975): "Lip and jaw motor control during speech", JSHR 19, 207-220.
- Moll, K.L., G.N. Zimmerman, and A. Smith (1977): "The study of speech production as a human neuromotor system", in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.), 107-127, Tokyo: University of Tokyo Press.

SYMPOSIUM NO. 7: THE RELATION BETWEEN SENTENCE PROSODY AND WORD PROSODY

(see vol. II, p. 375-430)

Moderator: Eva Gårding

Panelists: Arthur S. Abramson, Gösta Bruce, Johan 't Hart,  
Eunice V. Pike, Nina Thorsen, and Kay Williamson

Chairperson: George D. Allen

EVA GÅRDING'S INTRODUCTION

The purpose of the symposium is to discuss the relation between sentence prosody and word prosody in different prosodic systems, with the aim of tracking down universal features and tendencies in this relation. A more general goal is to contribute to a common framework for the description of prosodic phenomena. Since one of the symposia deals with length, such features have not been included here. To secure a broad treatment of the topic, a number of specialists of various prosodic systems were invited to be members of the panel. They represent Thai (Abramson), Amerindian languages (Pike), Nigerian languages (Williamson), Swedish (Bruce), Danish (Thorsen), Dutch ('t Hart), and Czech (Jánota).<sup>1</sup>

In volume II p.375 I proposed a terminology and suggested some points for discussion. I shall first elaborate on these points (1.1 - 1.4). Next follow summaries of the panelists' comments to their written contributions (2) and then an account of the discussion, ordered by subject (3.0 - 3.3). With this order some of the contributions have had to be split up under different headings. Finally I try to give a short evaluation of the symposium (4).

1.1 Basic units<sup>2</sup>

The first basic concept which is fundamental to our discussion is sentence intonation. Everybody on the panel agrees that an observed pitch pattern is equal to sentence intonation plus word intonation. But there are different views about what these two components really are and how they should be extracted from an observed curve. For those who treat tone languages and 2-accent languages, sentence intonation seems to be a broad general fea-

1) Přemysl Jánota was unable to attend the congress.

2) See footnote on page 293.

ture (called global in what follows), possibly combined with a local feature. These features express the illocutionary character of an utterance, for instance, statement or question. They can be manifested as downdrift or absence of downdrift with or without some consistent local glide. The ups and downs determined by the tones and accents are imposed on this pattern.

For 't Hart and Collier in their analysis of Dutch, however, intonation is the total intonation pattern including the rises and falls over the accents. Word prosody is lexical accentuation and it only determines the timing of some salient parts in the pattern. Palmer (1922), Bolinger (1958), and O'Connor and Arnold (1961) have described the intonation of English in a similar way.

It seems clear that the existence of these two radically different interpretations does not facilitate our task.

In connection with the concept sentence intonation we should perhaps ask ourselves the following questions:

Are the prosodic systems really so different that they have to be analysed differently?

Is a compromise possible so that sentence intonation can be given the same meaning in different prosodic systems?

Are there any languages for which the decomposition into word prosody and sentence prosody is meaningless?

Is there perhaps a need for a smaller unit between sentence and word, such as phrase?

The second concept important for our discussion is sentence accent. Even here there is fundamental disagreement. About half of the panel take sentence accent to be an accent feature expressing the focus of a sentence which can signal semantically or emotionally important words. In widely different prosodic systems, sentence accent has been reported to have similar manifestations: increased duration and amplitude in combination with a special pitch pattern. Most often sentence accent occurs on the accented syllable of the word in focus but it can also have a separate manifestation on a later syllable. Such cases have been reported by Eunice Pike for Ayutla Mixtec and Acatlan Mixtec (p.414) and by Gösta Bruce and myself for Swedish dialects (p.388). As a rule the tone languages listed by Eunice Pike have sentence accent. Kay Williamson, on the other hand, does not need the concept for her description of Nigerian tone languages and Nina Thorsen as-

cribes the prominent accents elicited from Copenhagen speakers to emphasis or contrast.

't Hart and Collier do not separate a special sentence accent from other accents. All pitch movements in combination with accented syllables are sentence accents. This is consistent with their view of intonation.

The sentence accent has been very useful in the analysis of Swedish intonation and I am ethnocentric enough to think that it should be useful generally. I therefore suggest that we discuss the relevance and usefulness of sentence accent. Also here we might need an intermediate level between word and sentence. A parallel term to phrase intonation would be phrase accent.

The other basic units are of course accents and tones but competing descriptions of tones and accents, although abundant in the literature,<sup>1</sup> are not to be found in the contributions to this symposium. They may come up in the open discussion, however.

#### 1.2 Extraction of the phonetic correlates of basic units

Suppose now that we have some idea of the linguistic nature of the basic prosodic units at sentence and word level. How should we extract their phonetic correlates from observed pitch patterns? To do this extraction it seems necessary to consider utterances in which sentence prosody and word prosody are varied in a systematic fashion. This is the method which has been used by Gösta Bruce. The method may lead to basic forms that are not always directly observable in a given pattern. For Swedish dialects we have in this way extracted four different manifestations of sentence accent which are extremely useful in generating and explaining the different types of intonation in Swedish dialects.

For Abramson it is the citation form which contains the phonetic correlates of the basic tone and this form is then perturbed by sentence prosody and adjacent tones.

There are hardly any competing views about the phonetic correlates of tones but for accents the pendulum has swung between pitch and intensity. For a long time now it has been customary to regard all accents as pitch accents. I found it very refreshing to see the data presented by Fujisaki and his collaborators in a poster session at this congress (Fujisaki et al., 1979a). The data seemed to reestablish some of the importance of intensity for English accents as compared to Japanese ones.

1) See e.g. references in Leben (1978).

For sentence intonation, various auxiliary lines have been proposed. 't Hart and his collaborators have used a baseline joining local minima in a curve, only for them it does not represent sentence intonation.<sup>1</sup> Nina Thorsen joins points (lows) representing stressed syllables. For Swedish we have used a more complex construction of baselines and topline (Bruce and Gårding, 1979). Common to all these constructions is a baseline whose steepness is determined by the length of the phrase. In Fujisaki's intonation model, which he showed during the discussion ensuing the report on perception, the baseline is independent of the length of the utterance (Fujisaki et al., 1979b). I have asked him to give a brief demonstration of the pertinent parts of his intonation model at the end of the time allotted to the panelists.

To sum up my questions under this point (1.2):

I suggest that we discuss various methods for the extraction of the phonetic correlates of the prosodic units.

How should this extraction be done and to what purpose?

Are principally different methods possible?

And what are the phonetic correlates of the basic units, sentence intonation, sentence accent, lexical tone, lexical accent?

### 1.3 Interaction between sentence prosody and word prosody

Let us now assume that we have extracted the phonetic correlates of the basic units of sentence prosody and word prosody. To generate perceptually correct pitch patterns we must know how these units interact. And here finally we come to the main theme.

Generally speaking, sentence prosody precedes and sets the scale for word prosody. This must be a true universal. For instance, downdrift influences everything on its way, and in Swedish, sentence accent influences all preceding and following word accents.

Apart from the interaction between sentence prosody and word prosody there is also interaction between adjacent units in the utterance, usually called tonal coarticulation and described by tone rules (Hyman and Schuh, 1974; Schuh, 1978).

I suggest the following points of discussion under 3:

Is the order sentence prosody, word prosody a true hierarchy?

And at the sentence level, is sentence intonation primary to sentence accent?

Are there any general principles governing tonal and accentual coarticulation?

1) 't Hart modifies this statement: The baseline is not the only manifestation of sentence intonation.

### 1.4 Additional questions

Here I collect questions which are marginal to the main theme. How does one determine if the basic prosodic unit for a word is a tone or an accent? According to Eunice Pike it is possible to determine if a given High represents an accent or a tone by studying its effect on vowel quality. Accented syllables have full vowels and unaccented vowels are reduced. Also accented consonants are affected. High tone, on the other hand, has no influence on vowel quality.

Accent also affects duration in a drastic way. In Swedish an accented syllable is more than twice as long as an unaccented one, whereas tone only has a marginal effect on duration.

According to many linguists, e.g. Larry Hyman (1975, p. 207 ff.) the difference between tone and accent is a linguistic one, not a phonetic one. I think that this point should be debated further. Tone and accent seem to have quite different contextual effects, difficult to explain without some difference of physiology.

## 2. COMMENTS FROM THE PANELISTS

Arthur Abramson emphasizes that the five tones of Thai are essentially preserved in connected speech.<sup>1</sup> He goes on to give an example which shows that the declination over an utterance is 30% of a woman's voice range, with the topline responsible for a larger amount of the declination than the baseline. Sentence accent is perhaps not as adequate a notion for the description of Thai as syntactic groupings in which phrase breaks are signalled by prosodic variation.

Eunice Pike summarizes ways in which pitch is used in the languages she has studied. It signals contrasts between lexical items, segments a stream of speech into words and clauses, marks sentence stress and conveys attitudinal meaning. Eunice Pike exemplifies these functions in various languages. In Marinahua of Peru a high tone will be still higher and a low tone lower under sentence stress. In Mikasuki of Florida tones are modified downward to mark boundaries between words and upward to mark bound-

1) According to Gsell (1979) the distinctiveness of tone in Thai is very much reduced in connected speech. There are only certain positions, comparable to accented syllables, in which the tones retain their distinctive power. - This publication contains a lot of other information relevant to the theme of this symposium.

aries between phrases. In Eastern Popoloc of Mexico a final upglide marks politeness as opposed to the unmarked neutral ending with a glottal stop. (For references see Vol. II p. 416). In Fasu high tone and low tone contrast lexical items only in stressed syllables, the unstressed syllables carry attitude or sentence intonation. A special voice quality is used in talk with spirits.

Kay Williamson calls attention to tonal modifications due to grammatical constructions which in her present view were underemphasized in her earlier contribution (p. 424). With fewer minimal pairs there is more freedom for extensive variation without causing ambiguity. One of the languages has some dialects which could be called pitch accent systems. Such a system may have developed as follows. Series of high tones have gone low and the surviving highs have become - phrase accents! Kay Williamson exemplifies global and local effects in connection with sentence type. Global manifestations are downdrift, a cancelling of downdrift or a raising of highs so as to increase intervals. One example of a local effect is that in Igbo the normal pronominal repetition of a subject at the beginning of a phrase has a high tone in the statement and a low tone in the question. In all other cases the local effect occurs at the end of the sentence with an opposition between statement and question. There is a final high for statement as opposed to low for question in some of the languages, which goes to show that the connection of high with question and low with statement is not a universal one.

Gösta Bruce shows a Stockholm Swedish pitch contour with six word accents surrounding a sentence accent in the middle of the utterance (Fig. 1). This figure shows that there are two contextual variants of one and the same accent, depending on their position relative to the sentence accent, rise-falls before the sentence accent and mere falls after it. Statement intonation is represented by the downdrift. The extent of this downdrift for a given speaker

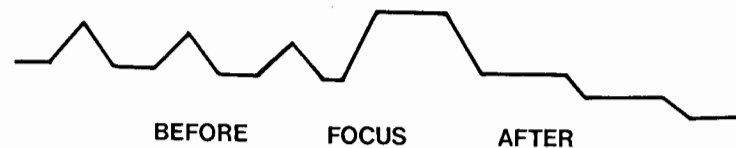


Fig. 1. Downdrift in Swedish. Stylized pitch contour of a Swedish utterance. From Gösta Bruce. Work in progress.

seems to be independent of the length of the utterance. However, the figure, assumed to be typical in this respect, shows that the actual course of the downdrift pattern has a very gentle slope before the sentence accent and a steeper, terrace-shaped downdrift afterwards. The figure sums up some important aspects of the interaction between sentence prosody and word prosody. Sentence intonation sets the scale for accentuation and accentuation determines the time course, in this case of the downdrift.

Nina Thorsen needs two prosodic units between word and sentence, the stress group, defined as the stressed syllable and the succession of unstressed ones, and a prosodic phrase group consisting of several stress groups. In her prosodic system there are two components which do not interact. Stress-group patterns are simply superimposed on the intonation contour which in her model is described as a line joining the stressed syllables. Nina Thorsen further discusses problems of definition when she applies this view to utterances with emphasis for contrast. She prefers to think that with emphasis the utterance is reduced tonally to a one-stress utterance. With this interpretation the difference between statement and question lies mainly in the stressed syllable and the post-tonic syllables.

Johan 't Hart underlines that in his and his collaborators' analysis of Dutch, declination is part of the intonation but not the only manifestation of it. Word prosody is lexical accentuation and sentence accentuation is represented by the pitch accents in the sentence. Sentence intonation has a higher place in the hierarchy. Reference to the communicative function has been avoided. Intonation patterns are not connected with linguistic categories such as statements, questions, wishes or commands, but represent classes of melodic shapes distinguished by the listener.

Hiroya Fujisaki in an extra contribution invited by the moderator, describes a model for Japanese intonation. It is, he says, principally similar to an intonation model proposed by Öhman (1967). In logarithmic scale all  $F_0$  patterns are sums of two components, a baseline component (called voicing component) corresponding to sentence prosody and an accent component. Fujisaki showed a figure (Fig. 2) that strengthens his view that the time constant of the baseline is not affected by sentence length. In longer sentences the speaker resets his baseline at one of the major syntactic boundaries. A general observation is that with an

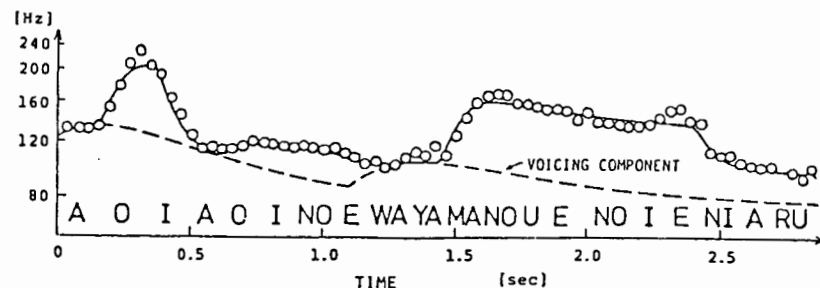


Fig. 2. Analysis by synthesis of a Japanese  $F_0$ -contour with two voicing (baseline) commands. From Fujisaki et al. (1979b).

absolute scale the height of accentual  $F_0$  peaks over the baseline decreases towards the end of a declarative intonation contour. In logarithmic scale, however, the peaks have approximately the same height over the baseline. This analysis can lead to a simpler and more illuminating interpretation of prosody.

### 3. DISCUSSION

In this section I have chosen to organize the discussion by subject. Consequently one intervention may occur in several places. I have followed the terminology of each discussant, inserting my earlier suggested term within parentheses. Terminological remarks, in particular those with a bearing on typology, have been collected under point 3.0. Since all the additional questions (1.4) concern the basic units and their correlates, they have been referred to 3.1 and 3.2. Otherwise the points for discussion follow the suggested outline. The discussion typically begins with the panel, proceeds with the respondents from the audience and ends with the panelists' responses.

#### 3.0 Terminology

Irmgard Mahnken wants the terminology to show the non-isomorphic character between grammatical and prosodic units.

William Moulton offers a list of terms useful for the description of different prosodic systems. Three uses of pitch and stress, lexical, morphological and syntactical, can be combined in different ways. William Moulton also underlines the need to distinguish between gradient versus discrete pitch and stress signals.

#### 3.1 Basic units

All the panelists agree on the usefulness of an intermediate unit between sentence and word level.

For the description of a Subject Object Verb language, Kay Williamson uses the concept tone group. This tone group is syntactically determined. Within such a group the first word sets the pattern for the whole group. For the group Object Verb the verb loses its own pattern and follows that of the object. In the dialects mentioned earlier, where only one High per group survives, normally the last one, group accent might be an appropriate term.

Also Johan 't Hart advocates the idea of introducing groups into the descriptive framework.

Eva Gårding argues that in the data presented by Arthur Abramson for Thai (p.383) one can find phrase accents manifested as increased amplitude and length and in the same utterance also something that looks like a sentence accent with an even more prominent increase of length and amplitude. In her own dialect of Swedish there are similar phenomena. Lexical restrictions on the pitch pattern in an accent language like Swedish make it perhaps more convenient to signal a syntactic unit by a phrase accent, expressed by increased amplitude and length rather than by a particular pitch configuration, as for instance in the Dutch hat pattern.

Arthur Abramson agrees with this interpretation of phrase accent in his material but he is not happy with the notion of sentence accent, which is determined by the whole discourse.

René Gsell gives a linguistic functional definition of tone, accent and sentence which he missed in the panelists' discussion. (This critique was repeated by other discussants, e.g. Mahnken, Moulton and Carton.) Tone is a paradigmatic mark of morphemes and words. Accent is a syntagmatic mark and the function of accent is the grouping of morphemes into words and at a higher level, of words into tagmemes and larger phrase constituents. In the symposium sentence accent has been used for emphasis and focus, which are two different things. From a linguistic point of view sentence accent is mainly phrase accent, the culminative mark of a higher constituent. Intonation is a still higher level of integration by which tagmemes or constituents are grouped into sentences.

Vichin Panupong demonstrates how in Thai sentence intonation can be signalled by final tone-bearing particles. One such particle is ka which modifies the total meaning of a sentence from e.g. statement to question by means of one of four possible tones. Sentence intonation can be carried by a final word as well. Final particles are also used to mark boundaries.



Sieb Nooteboom comments on the confusion between pitch accent in the Dutch analysis as compared to sentence accent in the Swedish one. The Swedish picture of one accent determined by focus surrounded by a number of smaller ripples caused by other accents (Fig.1) may correspond to just one pitch accent in Dutch determined by focus without any pitch manifestation of the other accents. Gösta Bruce has analysed sentences with only one semantically determined pitch accent whereas 't Hart (p.398) shows sentences with a number of semantically determined pitch accents. The question is what would happen in Swedish in a comparable situation, i.e. in a sentence with several semantically determined pitch accents.

Fernand Carton points out that even within one language there are problems of description. He needs the notion of accent (as do other analysts) for his study of dialects in the north of France where accent is still contrastive. Other analysts, as e.g. Mario Rossi, claim that there is no accent in modern French since it has only demarcative (syntactic) function. A common theoretical framework is needed, which takes functional aspects as well as the existence of different factors into account. A constant check on the interplay between form and substance is needed at all stages of the analysis and perceptual tests are crucial.

Alan Cruttenden is disturbed by the continued use of such simple categories as statements and questions for sentence intonation.

Barbara Frohovník thinks that an intermediate unit like prosodic phrase might have a bearing on the definition of the word and the sentence.

Lisa Selkirk with experience from comparative work in French and English wants to posit an intermediate level which has a syntactic definition.

Philippe Martin wonders how phoneticians can say that there are well formed sequences of pitch accents, as for instance in Dutch, if they reject any relation between syntax and sentence intonation.

#### Responses to 3.0 and 3.1

Gösta Bruce answers Sieb Nooteboom that there may be two or three sentence accents in the same Swedish utterance.

Eva Gårding is of the opinion that all panelists agree with René Gsell on the importance of function in a linguistic analysis.<sup>1</sup>

Kay Williamson in response to William Moulton's typological suggestions says that at least nine combinations of pitch and stress are needed. We speak of tone languages, stress languages and pitch accent languages, but we need more categories for the languages described in Eunice Pike's contribution, where both stress and pitch are contrastive. There are in addition at least two types of tone languages, the syllable-tone type and the word-tone type. To sum up, we need a rather more complex typology than the ones suggested earlier.

Eva Gårding reassures Alan Cruttenden that the members of the panel are well aware of the existence of a variety of sentence intonation types. The reason there is so much talk of statement and question intonation in the contributions is that the purpose of the symposium is to study the relation between word and sentence prosody and that this can be done safely in the statement and question types since they are well established in prosodic systems and easily elicited from speakers.

#### 3.2 Extraction of the phonetic correlates of basic units

##### 3.2.1. Citation forms versus other forms

According to Gösta Bruce citation forms would be insufficient for a thorough analysis of an accent language like Swedish. A Swedish citation form is a very complex pattern containing contributions from several linguistic variables, word accent, sentence accent, sentence intonation and terminal juncture. His results have been obtained by comparing words in different prosodic contexts. In this way it has been possible to decompose the classical double-peaked Accent 2 pattern of e.g. Stockholm Swedish into a word accent fall, a sentence accent rise and a terminal juncture fall.

Arthur Abramson defends the use of citation forms, partly for practical reasons - they are easy to elicit and measure - and partly for psychological reasons - children tend to learn one-word

---

1) I was too rash here. Gsell and Moulton and others requested a functional definition of the concepts under discussion. It should have been said from the beginning that the basic units were intended to be useful and efficient in the analysis and synthesis of prosody. In this capacity they are not necessarily functional units in the classical sense.



utterances and hence citation forms.

Alan Cruttenden gives an example from one variety of Panjabi which supports the view that the basic form of pitch accent should be derived from connected speech rather than citation forms. In connected speech a two-way pitch accent distinction involves a clear deviation downwards or upwards respectively in a particular intonation pattern, whereas in citation forms the distinction is very complex.

Eunice Pike finds it very important to remember in an aural linguistic analysis that lexical tones may be modified by sentence intonation or sentence stress. One trick in such an analysis is to ask for three items and have the words you want to contrast as number one and two. These two will then have a chance to have the same intonation pattern whereas the last item will have terminal intonation. To separate sentence accent from lexical tone it is advisable to have at least two words in a sequence. One of these words will then have the sentence accent and the other words will carry only tone.

### 3.2.2. Methods for the extraction of basic forms

At least four methods have been mentioned in the contributions, elicitation of citation forms (Abramson), comparison of prosodic variables in different contexts (Bruce, Pike), analysis by perception ('t Hart), and analysis by synthesis (Fujisaki).

Edward Purcell makes a request for more statistically based approaches to modelling tone and intonation, by using e.g. polynomial regression. It might then be possible to solve equivalence problems like the Dutch and Swedish sentence accent.

Yukihiko Nishinuma points out that an intonation model has to take the integration of independent acoustic parameters into account as well as the effect of masking at different levels.

### Responses to 3.2.2

Johan 't Hart argues that the most important need is not statistics but a large inventory of intonational possibilities and perceptual testing. He would like to know if Hiroya Fujisaki is as concerned about the fit between synthetic and perceptual patterns as he is about the fit between synthetic and acoustic ones. As for logarithmic versus linear scale he does not think it matters much in short utterances.

Arthur Abramson is in sympathy with the use of polynomial regression but finds it most often sufficient to form hypotheses

based on the acoustic manifestations and to test these hypotheses perceptually.

### 3.2.3 Phonetic correlates

#### a) Sentence intonation and downdrift

Nina Thorsen points out that in her Danish material downdrift is evenly distributed over the utterance. The downdrift does not occur only in connection with the accented syllables as shown in Bruce's figure (Fig. 1). Also, the range varies with the length of a sentence within certain limits. Contrary to Fujisaki's model for Japanese, the downdrift in her material is a linear function of the length of a short utterance. In long ones there is a resetting of intonation in connection with syntactic boundaries. She referred to the figure (Vol. II p. 417) where it appears that the height of the post-tonic syllables above the "baseline" does decrease toward the end, even with a logarithmic scale.

Gösta Bruce ascribes the difference between the distribution of downdrift in Swedish and Danish to the different use of sentence accent. In standard Swedish a normal neutral utterance will have sentence accent on the last accented word whereas in Danish and perhaps also in Southern Swedish dialects there is no obligatory rule. The range of the downdrift has appeared to be constant in sentences with two, three and four accented syllables.

Osamu Fujimura mentions work on pitch synthesis conducted by Janet Pierrehumbert at Bell Laboratories. It is somewhat similar to the work reported by Hiroya Fujisaki. The algorithm is based on specifications of pitch peaks representing relative prominence with options for low-tone stress. Nuclear tones fall below the baseline and postnuclear tones are neutralized. Pitch declination is a descending time function with resetting at major phrase boundaries (see Pierrehumbert, 1979).

Hiroya Fujisaki agrees with Johan 't Hart that the scale is not so important within a small range but for longer sentences the distinction is very clear. In answer to Nina Thorsen he says that there may be many language-specific points in prosody. He strongly agrees with Edward Purcell about the need for analytic and quantitative methods in the analysis of the production and perception of prosodic phenomena.

#### b) Accent versus tone and accent versus stress

In her description of the dialects of İzön Kay Williamson tries to show that there is a gliding scale between tone-dialects

and accent-dialects with a very narrow cross-over zone. In general, [and this is consistent with Eunice Pike's description, EG] the more you have a tone language, the more things are symmetrical, and the more you have an accent language, the less things are symmetrical. The accents have more prominence and other things get reduced in relation to it. Perhaps this is the reason why it is easier to talk about sentence accent in accent languages than in tone languages.

René Gsell says that from a functional point of view the Scandinavian languages, even Danish, are tone languages. The 'stød' acts as an intonation depressor and is a clear example of interaction between word and sentence prosody.

Yukihiro Nishinuma (and also Irmgard Mahnken) find that in the discussion of intonation too much emphasis is put on  $F_0$ , although everybody who has worked on automatic intonation detection knows that  $F_0$  is not sufficient.

Ivan Fónagy presents the acoustic correlates of a Hungarian phrase akar, a kar (with accent on the first and second syllable respectively) as a statement and as a question in normally intoned and whispered speech, by which he wants to show that pitch accent is not an appropriate term for the acoustic correlates of the accent. As a term he prefers stress.

#### Responses to 3.2.3

Eva Gårding agrees with the view that too much emphasis has been put on  $F_0$ . This trend seems to have been weakened lately.

Arthur Abramson points out that apart from fundamental frequency and amplitude variations there are also other cues that may have signal value, creaky voice and various other forms of laryngeal constriction.

### 3.3 Interaction between sentence prosody and word prosody

#### 3.3.1 Hierarchy

Three views are represented at the symposium: Sentence prosody is primary (e.g. Bruce, 't Hart), lexical prosody is primary (Abramson), and sentence prosody and lexical prosody are at the same level. The last view is implied by the model presented by Hiroya Fujisaki. Here the word-prosodic part and the sentence-prosodic part are extracted simultaneously from an observed curve and may therefore be regarded as belonging to the same level of the hierarchy. The final  $F_0$  contour is the sum of these two parts.

Arthur Abramson's feeling is that lexical prosody must be paramount in a tone language. In the mental lexicon the storage form must carry the tone as part of the morpheme. When these tones are strung together in connected speech a particular intonation is superimposed.

According to Johan 't Hart there is a higher hierarchical position for sentence intonation.

René Gsell claims that with the definitions he has given earlier (see 3.2) it is easier to understand interaction. At each level a higher constituent mark modifies lower constituent marks. Intonation dominates sentence accents, sentence accents dominate the word accents and so on. The phonetic characteristics of lower marks are not indifferent to the grouping of higher layers.

Einar Haugen remarks that the Scandinavian word accents are part of the stress pattern of the sentence and always to be seen in relation to the whole utterance. Therefore, to ask whether the word or the utterance is primary is a chicken-and-egg kind of question. You cannot say any Swedish or Norwegian word without having both tone and sentence intonation. They are stored with the word. Every native knows which tone a word has, although it never occurs without sentence intonation. Accent 2 has to be interpreted as a perturbation of the unmarked sentence intonation.

#### Responses to 3.3.1

Eva Gårding refers the conflicting views about the hierarchical relation between sentence prosody and word prosody to different points of departure. To work out a program for pitch synthesis by rule you need a rough idea of the sentence intonation, i.e., where to start on the frequency scale etc. So with this aim in view it is very natural to regard sentence intonation as primary. But with a psycholinguistic approach you are interested in the forms stored in the memory and the citation forms become primary in your hierarchy. These will then be perturbed by sentence prosody at some secondary level, the phrase or the sentence level.<sup>1</sup>

#### 3.3.2 Contextual interaction

Arthur Abramson points out that sandhi phenomena are phonological and have nothing to do with the interaction treated in this section.

1) Gabrielle Konopczynski suggests in a written contribution submitted after the symposium that one should look for a hierarchy by studying in detail how children acquire tone languages.

Gösta Bruce's figure (Fig. 1) gives a good example of interaction between sentence accent and word accents on the one hand and sentence accent and sentence intonation on the other.

George Allen is interested in the deletion of postnuclear accented syllables in an English phrase. This pattern seems to be acquired quite early by children, at the age of 30 to 36 months.

Osamu Fujimura remarks that problems of accentual patterns, such as interaction between sentence accent and lexical accents have been discussed extensively in the traditional linguistic literature in Japanese. He wants to call attention to McCawley's (1968) monograph.

Perceptual tests have shown that the pitch declination effect is compensated for by listeners when they judge the height of accent peaks (Pierrehumbert, 1979).

Ana Tataru exemplifies different relations between word accent and sentence accent in Romanian on the one hand and English and German on the other. Such differences are of great pedagogical interest.<sup>1</sup>

### 3.3.3 Word prosody restricting sentence prosody

Gösta Bruce comments on the often heard assumption that a speaker of an accent language like Swedish is less free in his or her use of pitch as an expression of sentence type and attitude than a speaker of another language, like Dutch for instance. There are restrictions in the possible use of pitch movements locally but globally you are free to express other aspects of intonation.

Johan 't Hart points out that in Dutch there are also restrictions. After a rise, pitch has to come down again to be ready for the next rise. He refers to the examples given in his contribution (p.398) to show that there are also restrictions in the placement of the pitch movement which may to some extent be determined by the syntactic boundaries.

Einar Haugen reminds the audience of Otto Jespersen, who claimed that Norwegians and Swedes were unable to express nuances of feeling as well as Danes, because of the tones. It was to disprove this point that Einar Haugen went into the study of tone!

---

1) Paul Schäfersküpper in a written contribution points out that in German, sentence accent operates over larger domains than the syllable.

### 4. MODERATOR'S AFTERTHOUGHTS

The aim of the symposium was to discuss word prosody and sentence prosody and the relation between them. Although precise results or general agreement were not to be expected, the symposium has contributed new material and well-taken points, and it has put some important questions into focus. I shall list some of them here.

It seems that even a large number of prosodic systems, as varied as those represented at the symposium, are sufficiently similar to be treated in a common framework, and that the dichotomy between word prosody, which I would now prefer to call lexical prosody, and sentence prosody, including phrase prosody, is useful even in languages whose lexical prosody is predictable from simple rules.

To find the basic units of the dichotomy we need data from all levels of analysis on which models can be based. I especially want to stress the need for simple but strict generative models. These models should aim at simulating observed patterns of pitch ( $F_0$ ), intensity and duration. Without such models the interaction between word prosody and sentence prosody cannot be stated with a sufficient degree of precision.

The symposium has given strong evidence for some general tendencies in the interaction between sentence prosody and word prosody. Declination or downdrift has been observed for many languages representing a variety of prosodic systems. We have seen in the Swedish material how this gradual downdrift may be checked by an intervening sentence accent (Fig.1). It is quite possible that there are phonological systems where downdrift is masked by a late obligatory sentence or phrase accent.

Accent reduction brings out an interesting tendency. After the sentence accent (nuclear stress) all following accents tend to be reduced. There is evidence for this from Danish, Dutch, Swedish and Japanese (see Fujimura's intervention). This may be one of the asymmetries that Kay Williamson and Eunice Pike found typical of an accent language as compared to a tone language. A worthwhile project would be to explore the physiological background of this effect.

It has often been observed that the heights of equally strong accents decrease over a declining baseline. As pointed out by

Hiroya Fujisaki, however, their absolute heights are proportional to that of the baseline. This may be a universal.

Are there any general principles behind tonal and accentual coarticulation? This question was left unanswered. One of the reasons may be that these relations can only be studied together with durational aspects which were not included in the topics of the symposium.

#### References

- Bolinger, D.L. (1958): "A theory of pitch accent in English", Word 14, 109-149.
- Bruce, G. and E. Gårding (1978): "A prosodic typology for Swedish dialects", in Nordic prosody, E. Gårding, G. Bruce, and R. Bannert (eds.), 219-228, Travaux de l'Institut de linguistique de Lund 13.
- Fujisaki, H., K. Hirose, and M. Sugitō (1979a): "Comparison of word accent features in English and Japanese", Proc.Phon. 9, 376, Copenhagen: Institute of Phonetics.
- Fujisaki, H., K. Hirose, and K. Ohta (1979b): "Acoustic features of the fundamental frequency contours of declarative sentences in Japanese", RILP 13, 163-173.
- Gsell, R. (1979): Sur la prosodie du Thai standard: Tons et accent, Paris: Université de la Sorbonne Nouvelle.
- Hyman, L. (1975): Phonology: theory and analysis, New York: Holt, Rinehart and Winston.
- Hyman, L. and R.G. Schuh (1974): "Universals of tone rules: Evidence from West Africa", Linguistic Inquiry 5, 81-115.
- Leben, W.R. (1978): "The representation of tone", in Tone, V. Fromkin (ed.), 177-219, New York: Academic Press.
- McCawley, J.D. (1968): The phonological component of a grammar of Japanese, The Hague: Mouton.
- O'Connor, J.D. and G.F. Arnold (1961): Intonation of colloquial English, London: Longmans.
- Öhman, S.E.G. (1967): "Word and sentence intonation: A quantitative model", STL-QPSR 2-3, 20-54.
- Palmer, H.E. (1922): English intonation, Cambridge: Heffer.
- Pierrehumbert, J. (1979): "The perception of fundamental frequency declination", JASA 66, 363-369.
- Schuh, R.G. (1978): "Tone rules", in Tone, V. Fromkin (ed.), 221-256, New York: Academic Press.

## SYMPOSIUM NO. 8: THE PERCEPTION OF SPEECH VERSUS NONSPEECH

(see vol. II, p. 431-489)

Moderator: David B. Pisoni<sup>1</sup>

Panelists: Anthony E. Ades, Pierre L. Divenyi, Michael F. Dorman,  
Dominic W. Massaro, and Quentin Summerfield

Chairperson: Arthur S. Abramson

## DAVID B. PISONI's INTRODUCTION

Historically, the study of speech perception may be said to differ in a number of ways from the study of other aspects of auditory perception. First, the signals used to study the functioning of the auditory system were simple and discrete, typically varying along only a single physical dimension. By contrast, speech signals display very complex spectral and temporal relations. Although speech signals have also been varied along single physical dimensions, the perceptual consequences of such manipulation have not always followed from "equivalent" stimulations of a nonspeech nature. Alternatively, we may presume that the complexity of the spectral and temporal structure of speech and its variation is one additional source of perceptual differences between speech and nonspeech signals. Second, most of the research dealing with auditory psychophysics over the last thirty years has been concerned with the discriminative capacities of the sensory transducer and the functioning of the peripheral auditory mechanism. In the case of speech perception, however, the relevant mechanisms are assumed to be centrally located and intimately related to the more general cognitive processes that involve the encoding, storage and retrieval of information in memory. Moreover, experiments in auditory psychophysics have typically focused on experimental tasks and paradigms that involve discrimination rather than identification or recognition, processes thought to be most relevant to speech perception. All in all, it is generally believed that a good deal of what has been learned from research in auditory psychophysics and general auditory perception is only marginally relevant to the

---

1) David Pisoni could not be present at the congress and Michael Studdert-Kennedy acted as moderator at the meeting. David Pisoni is author of the introduction below.

study of speech perception and to an understanding of the underlying perceptual mechanisms. This situation has changed for the better in recent years as shown by the work of Dr. Divenyi and other psychophysicists who have become concerned with questions of speech perception. Despite these obvious differences, investigators have been interested in the differences in perception between speech and nonspeech signals. That such additional differences might exist was first suggested by the report of the earliest findings of categorical discrimination of speech by Liberman and his colleagues (1957). And it was with this general goal in mind that the first so-called "nonspeech control" experiment was carried out by Liberman and his colleagues (1961) in order to determine the basis for the apparent distinctiveness of speech sounds. In this study the spectrographic patterns for the /do/ and /to/ continuum were inverted producing a set of nonspeech patterns that differed in the onset time of the individual components. The results of perceptual tests showed peaks in discrimination for the speech stimuli replicating earlier findings. However, there was no evidence of comparable discrimination peaks for the nonspeech stimuli, a result that was interpreted at the time as further evidence for the distinctiveness of speech sounds and the effects of learning on speech perception. Numerous speech-nonspeech comparisons have been carried out over the years since these early studies, including several of the contributions to the present symposium. For the most part, these experiments have revealed results quite similar to the original findings of Liberman et al. Until quite recently, research reports have confirmed that performance with nonspeech control signals failed to show the same discrimination functions that were observed with the parallel set of speech signals (Cutting and Rosner, 1974; Miller et al., 1976; Pisoni, 1977). Subjects typically responded to the nonspeech signals at levels approximating chance performance. In more recent years, such differences in perception have been assumed to reflect two basically different modes of perception--a "speech mode" and an "auditory mode". Despite attempts to dismiss this dichotomy, additional evidence continues to accumulate as has been suggested by several of the new findings summarized in the papers included in this symposium.

The picture is far from clear, however, because the problems inherent in comparing speech and nonspeech signals have generated several questions about the interpretation of results obtained in earlier studies. First, there is the question of whether the same psychophysical properties found in the speech stimuli were really preserved in the parallel set of nonspeech control signals. Such a criticism is appropriate for the original /do/--/to/ nonspeech control stimuli which were simply inverted patterns reproduced on the pattern playback. The same remarks also apply to the well-known "chirp" and "bleat" control stimuli of Mattingly et al. (1971) which were created by removing the formant transitions and steady-states from the original speech context. These stimuli were presented in isolation to subjects for discrimination. Such manipulations, while nominally preserving the phonetic "cue" obviously result in marked changes in the spectral context of the signal which no doubt affects the detection and discrimination of the original formant transition. Such criticisms have been taken into account in the more recent experiments comparing speech and nonspeech signals as summarized by Dr. Dorman and Dr. Liberman, in which the stimulus materials remain identical across different experimental manipulations. While these more recent studies relieve some of the ambiguities of the earlier experiments, problems still remain in drawing comparisons between speech and nonspeech signals. For example, subjects in these experiments rarely practice with the nonspeech control signals to develop the competence required to categorize them consistently. With complex multi-dimensional signals it is quite difficult for subjects to attend to the relevant attributes that distinguish one signal from others presented in the experiment. A subject's performance with these nonspeech signals may therefore be no better than chance if he/she is not attending selectively to the same specific criterial attributes that distinguished the original speech stimuli. Indeed, not knowing what to listen for may force a subject to attend selectively to an irrelevant or misleading attribute of the signal itself. Alternatively, a subject may simply focus on the most salient auditory quality of the perceived stimulus without regard for the less salient acoustic properties which often are the most important in speech such as burst spectra or formant transitions. Since almost all of the nonspeech experiments conducted in the past were

carried out without the use of discrimination training and feedback to subjects, an observer may simply focus on one aspect of the stimulus on one trial and an entirely different aspect of the stimulus on the next trial. Without training experience to help the subject identify the criterial properties, the observed performance may be close to chance, a result that has been reported quite consistently in the literature. Setting aside some of these criticisms, the question still remains whether drawing comparisons in perception between speech and nonspeech signals will yield meaningful insights into the perceptual mechanisms deployed in processing speech. In recent years, the use of cross-language, developmental and comparative (i.e., cross-species) designs in speech perception research has proven to be quite useful in this regard as a way of separating out the various roles that genetic predispositions and experience play in speech perception. On the other hand, these types of investigations provide needed information about the course of learning and perceptual development since spoken language must be acquired in the local environment through social contact. On the other hand, comparative studies with both speech and nonspeech stimuli are useful in defining the lower limits on auditory system function. However, there are serious limitations in studies of this kind. For example, while it is cited with increasing frequency that chinchillas categorize synthetic stimuli differing in VOT in a manner quite similar to English-speaking adults, little if anything is ever mentioned, however, about the chinchilla's failure to carry out the same task with stimuli differing in the cues to place of articulation in stops, a discrimination that even young prelinguistic infants can make (Eimas, 1974). Should we then conclude that the English voicing contrast is purely sensory in origin, while place of articulation or voicing in Thai is somehow more "linguistic", brought on by inheritance or very early experience? With a little reflection, I think the answer must surely be negative. Such comparative studies are useful in speech perception research but only to the extent that they can specify the lower-limits on the sensory properties of the stimuli themselves. However, these findings are incapable, in principle, of providing any further information about how these signals might be "interpreted" or coded within the context of the experience and history of the organism.

Animals simply do not have spoken language and they do not and cannot recognize, as far as I know, differences between phonetic and phonological structure, a fundamental dichotomy in all natural languages. Cross-language and developmental designs have also been quite useful in providing new information about the role of early experience in perceptual development and the manner in which selective modification or tuning of the perceptual system takes place. Although the linguistic experience and background of a listener was once thought to control his/her discriminative capacities in speech perception experiments, recent findings strongly suggest that the perceptual system has a good deal of plasticity for retuning and realignment, even into adulthood. The extent to which control over the productive abilities remains plastic is still a topic to be explored. To what extent is it then useful to argue for the existence of different modes of perception for speech and nonspeech signals? Some investigators such as Dr. Ades would simply dismiss the distinctions drawn from earlier work on the grounds of parsimony and generality. He has argued recently (Ades, 1977) and in his contribution to this symposium that differences in perception between speech and nonspeech or consonants and vowels can be accounted for simply by recourse to the notion of "range" or the width of the context expressed in terms of the number of JNDs. As long as the range is small, absolute identification performance will be as good as differential discrimination. When the range is large, however, discrimination will be better than identification. Thus according to the account offered by Ades, a consonant continuum should display a smaller range than a vowel continuum. But as shown in Fig. 1 the facts are quite the reverse of his predictions.

In this figure we have reproduced the identification data collected by Perey and Pisoni (1977) in a magnitude estimation task. On each trial subjects had to respond to a stimulus with a rating on a scale from 1 to 7. One group of subjects received a consonant continuum differing in VOT, another received a vowel continuum. Through various transformations of the obtained stimulus-response matrix, scale scores were derived and an estimate of the perceived psychological spacing between stimuli was obtained. Scale scores are expressed in this figure in terms of



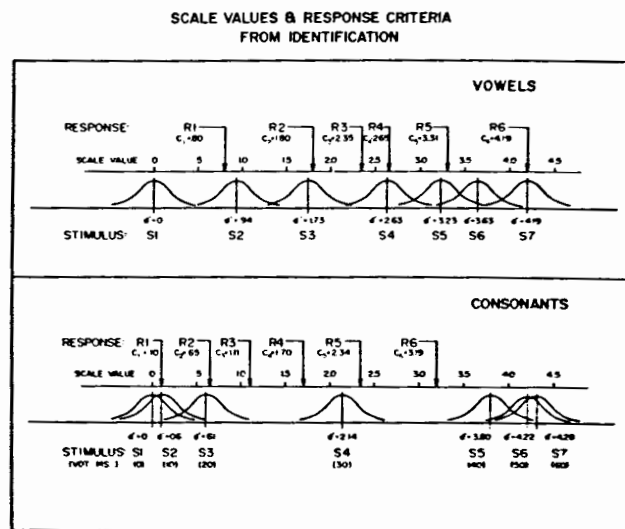


Figure 1. Scale values showing the perceived psychological space for consonants and vowels. Data were taken from Perey and Pisoni (1977) who required subjects to use a rating response in identification.

d's and by summing these individual values, an estimate of the total range or spacing of the stimuli was obtained. The cumulative  $d'$  is shown on the far right of each panel. Notice that the cumulative  $d'$  for the vowels shown on the top is 4.19 while the value for the consonants shown on the bottom is 4.28. If stimulus range were the correct explanation of the differences in perception between consonants and vowels as Dr. Ades would have us believe, the consonants should have displayed the smaller range. Obviously, this is simply not the case. However, what is of interest in this figure is the psychological spacing of signals within each panel. For the consonants, the spacing between adjacent stimuli is clearly unequal with a grouping close to the endpoints of the series. For the vowels, the spacing is more nearly equal across all the test stimuli suggesting the possibility of better resolution in discrimination, a result that has been known for

many years. Thus, Dr. Ades' argument that the range of stimuli can account for differences in perception between consonants and vowels or speech and nonspeech would seem to be incorrect, despite his attempts to generalize the Durlach and Braida (1969) model to speech perception. Moreover, this is a curious position to maintain anyway as it is commonly recognized, not only in speech perception research but in other areas of perceptual psychology, that "nominal" stimuli may receive differential amounts of processing or attention by the subject, that subjects may organize the interpretation of the sensory information differently under different conditions and that the sensory trace of the initial input signal may show only a faint resemblance to its final internal representation resulting from encoding and storage in memory. It is hard to deny that a speech signal elicits a characteristic mode of response in a human subject--a response that is not simply the consequence of an acoustic waveform leaving a meaningless sensory trace in the auditory periphery. Nevertheless, there is a great deal to learn about how the auditory system codes complex acoustic signals such as speech. Dr. Dorman, in summarizing work on the perception of transitions in speech and nonspeech context, has tried to establish the need for a specialized speech processor to account for differences in labelling of sine-wave stimuli when heard as either speech or nonspeech. Such explanations seem to me entirely premature at this time as the relevant psychophysical experiments with nonspeech signals have simply not been carried out yet. To remedy this state of affairs we have begun to collect labelling data in our laboratory recently using brief FM stimuli followed by a constant frequency (CF) steady-state. Schematized spectrograms of the test stimuli are shown in Fig. 2.

The left panel of this figure shows an idealized set of stimuli differing in the initial starting frequency of the FM. Three steady-state (CF) frequencies were selected, 850, 1500 and 2300 Hz. For each set we generated 21 test signals which spanned a range of 500 Hz above and below the CF of the steady-state component. In Experiment I the three sets of signals consisted of an isolated single component as shown on the left. In Experiment II we added an additional 500 Hz component to each of the original three sets of stimuli. Subjects were required to identify the



## FM TEST STIMULI

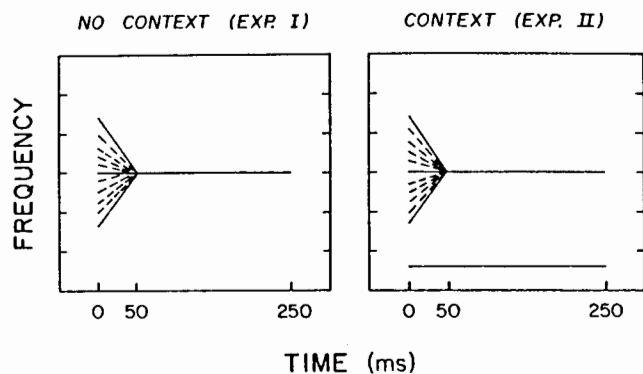


Figure 2. Schematized patterns showing the time course of the non-speech FM stimuli: The panel on the left illustrates the test stimuli without spectral context, the panel on the right shows the addition of a low frequency component to the same signals.

stimuli as "rising", "level" or "falling" after a brief training period with good exemplars selected from each category. The results of both experiments are shown in Fig. 3.

The labelling functions shown at the top for the three CF conditions reveal that the middle or "level" category response increases slightly in size as the CF of the steady-state increases from 850 Hz to 1500 Hz, a result that is consistent with what is known about frequency resolution in the auditory system. Over a wide range of frequencies, discrimination follows Weber's law. Thus, the level category should widen as the frequency of the steady-state increases for the same difference in initial starting frequency. Note that we have plotted starting frequency on a linear rather than log scale. The results for Experiment II in which an additional steady-state component was added are shown in the lower panel of the figure. Notice that for the 850 Hz condition the "level" category is now slightly larger than in the top panel suggesting the strong possibility of some interaction between the individual components. However, the other two condi-

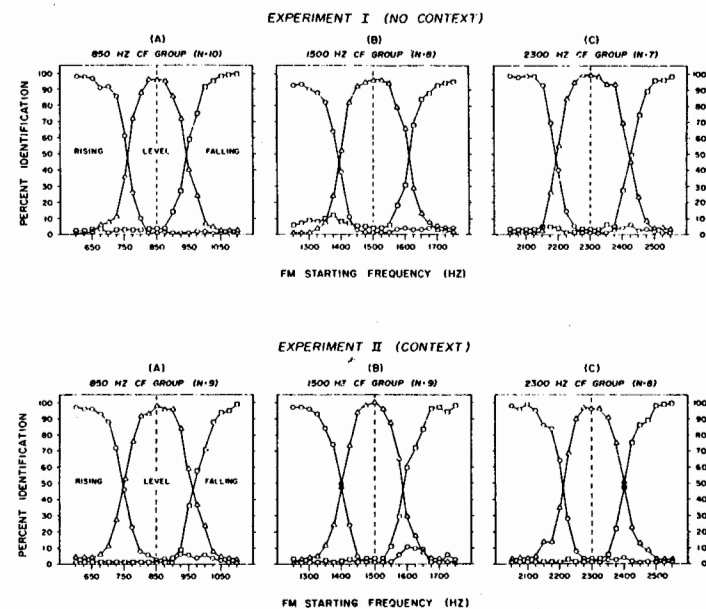


Figure 3. Identification data for FM stimuli obtained with three different steady-state CF's, 850 Hz, 1500 Hz and 2300 Hz. The top panel shows the identification data collected for FM's without context, the lower panel shows the data for test signals with the additional steady-state context present.

tions in Experiment II show a somewhat narrower range for the "level" category compared to the top panel indicating better resolution of frequency in the presence of another signal, a well known fact in auditory psychophysics. These recent findings were not originally intended to refute the arguments of Dorman and his colleagues who favor the postulation of some specialized perceptual mechanism for processing speech signals. Rather, I simply wanted to illustrate by way of example that the location of perceptual categories observed with nonspeech signals is not rigidly controlled by some simple physically defined invariant such as the direction of the frequency change. Moreover, as Dr. Divenyi has pointed out so well in his paper, we need to know much more about how the

basic constraints of the auditory system affect the way speech is initially coded for subsequent processing. Thus, in the present case several basic facts about frequency discrimination are sufficient to account for changes in our subjects' perceptual categorization of nonspeech FM's that are similar to speech. Whether it will be possible to generalize such psychophysical explanations to more complex signals such as speech remains to be seen from future research currently in progress in our laboratory and elsewhere.

In summary, there still appears to be good evidence for distinguishing between speech and nonspeech signals and for recognizing the existence of two distinct modes of perception, one associated with the sensory or psychophysical correlates of acoustic signals and the other with the interpretation and coding of acoustic signals as speech. Recent work has attempted to make these differences more precise by subjecting them to experimental test and searching for common underlying explanations. Taken together such results suggest to me that, just as in the case of "species-typical responding" observed in the behavior of other animals, the notion of a "speech mode" of perception captures certain aspects of the way human observers typically respond to speech signals that are highly familiar to them. We still do not know if it is simply a matter of familiarity as with music or whether there is something deeper and more closely related to biological survival of the organism. Nevertheless, such a conceptualization does not, at least in my view, commit one to the view that human listeners cannot respond to speech signals in other ways more closely correlated with the sensory or psychophysical attributes of the signals themselves. To deny the speech mode, however, is to ignore the fact that acoustic signals generated by the human vocal tract are used in a distinctive and quite systematic way by both talkers and listeners to communicate linguistically, a species-typical behavior that is restricted, as far as I know, to Homo sapiens.

Past experiments comparing the perception of speech and nonspeech signals have been quite useful in characterizing how the phonological systems of natural languages have, in some sense, made use of the general properties of sensory systems in selecting an inventory of phonetic features and their acoustic correlates (Stevens, 1972). The relatively small number of distinctive fea-

tures and their acoustic correlates that can be observed across a wide variety of diverse languages implies that there is a common sensory basis for language perception, a common means of controlling the mechanisms of speech production and a common cognitive definition of linguistic structure. Whether these facts are causally related will no doubt be a matter of much debate, speculation and new research in the years to come. It is clear, nevertheless, that the distinctions drawn in perception between speech and nonspeech signals still remain fundamental, setting apart research on speech perception from the study of auditory psychophysics and the field of auditory perception more generally.

#### Acknowledgements

The preparation of this paper was supported, in part, by NIMH grant MH-24027 and NINCDS grant NS-12179 to Indiana University in Bloomington. I am grateful to Peter Jusczyk and Jim Sawusch for comments on an earlier draft of the paper. Robert Remez discussed many of the theoretical issues summarized in the paper with me at length and provided helpful editorial comments that improved the overall exposition and quality. His help is greatly appreciated.

#### References

- Ades, A.E. (1977): "Vowels, consonants, speech and nonspeech", Psych. Rev. 84, 524-530.
- Cutting, J.E. and B.S. Rosner (1974): "Categories and boundaries in speech and music", Perc. Psych. 16, 564-570.
- Durlach, N.I. and L.D. Braida (1969): "Intensity perception I. Preliminary theory of intensity resolution", JASA 46, 372-383.
- Eimas, P.D. (1974): "Auditory and linguistic processing of cues for place of articulation by infants", Perc. Psych. 16, 513-521.
- Lieberman, A.M., K.S. Harris, H.S. Hoffman, and B.C. Griffith (1957): "The discrimination of speech sounds within and across phoneme boundaries", J.Exp.Psych. 54, 358-368.
- Lieberman, A.M., K.S. Harris, J.A. Kinney, and H.L. Lane (1961): "The discrimination of relative onset time of the components of certain speech and non-speech patterns", J.Exp.Psych. 61, 379-388.
- Mattingly, I.G., A.M. Liberman, A.K. Syrdal, and T.G. Halwes (1971): "Discrimination in speech and non-speech modes", Cogn.Psych. 2, 131-157.
- Miller, J.D., C.C. Wier, R. Pastore, W.J. Kelly, and R.J. Dooling (1976): "Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception", JASA 60, 410-417.
- Perey, A.J. and D.B. Pisoni (1977): "Dual processing versus response-limitation accounts of categorical perception: A reply to MacMillan, Kaplan and Creelman", JASA 62, S1, 60-61.

Pisoni, D.B. (1977): "Identification and discrimination of the relative onset of two component tones: Implications for voicing perception in stops", *JASA* 61, 1352-1361.

Stevens, K.N. (1972): "The quantal theory of speech: Evidence from articulatory-acoustic data", in *Human communication: A unified view*, E.E. David, Jr. and P.B. Denes (eds.), New York: McGraw-Hill.

#### COMMENTS FROM THE PANELISTS

The symposium on the perception of speech and nonspeech began with a brief summary statement by each of the contributors. This was followed by a panel discussion dealing with several issues that came up during the presentations. Finally, a number of questions and comments from the general audience were presented, followed by further discussion by the members of the panel. The highlights of these discussions and interactions are summarized below in an attempt to capture the flavor of the general issues and problems that surfaced as a result of this symposium.

Dr. Ades began his presentation by summarizing his paper contributed to the symposium and offering several comments on the introductory remarks given earlier by Professor Pisoni. Dr. Ades reiterated several times in this presentation that he personally believed that speech perception was, in some sense, "unique" or "special" despite the weak evidence usually cited from identification and discrimination experiments. He argued that the differences in perception between speech and nonspeech signals or consonants and vowels could be accounted for by differences in the range or spacing of signals. Dr. Ades criticized the recent data presented by Professor Pisoni showing equivalent ranges for consonants and vowels on the grounds that these data were collected in an identification rather than a discrimination paradigm. Most of Dr. Ades' specific remarks were directed, however, at narrow experimental questions, particularly the use of high uncertainty discrimination paradigms which provide relatively low estimates of discriminability.

Dr. Divenyi argued for the operation of two stages of processing in auditory perception regardless of whether the signals are complex auditory patterns or speech signals. According to Dr. Divenyi, speech is simply one class of complex signals with which the listener has had extensive experience and familiarity. Dr.

Divenyi described his two-stage model of auditory processing. The first stage, the auditory stage, involves the sensory analysis and coding of signals by the peripheral auditory system. The representation of signals at this stage is something like a neurogram reflecting the frequency selectivity of the auditory system. The second stage, the temporal stage in Dr. Divenyi's model, involves the analysis and coding of temporal information or patterns in both speech and nonspeech signals. Dr. Divenyi argued that the differences in perception between speech and nonspeech signals were due to differences in listening strategies brought about by learning and experience with speech and other sounds. Thus, in listening to speech several different strategies are available to the listener for centering or positioning the listening band differently. Dr. Divenyi concluded that there were no structural differences in perception between the so-called "speech mode" and "nonspeech modes" of processing. The distinctiveness of speech arises, according to Dr. Divenyi, from mere exposure and familiarity with speech and not because of any specialized processing by the auditory system.

Professor Dorman summarized his recent research which was carried out in collaboration with Drs. Bailey and Summerfield. This work was concerned with the perception of speech and nonspeech stimuli differing in the cues to place of articulation. Professor Dorman stated that his interest in these comparisons grew out of several questions surrounding whether infants can perceive speech signals as speech rather than simply complex nonspeech patterns. The methodology employed in these studies using adult subjects involved comparisons dissociating the location of the "phonetic" boundary from the location of the "acoustic" boundary. The results of these tests showed differences in the loci of the boundaries depending on whether the nonspeech stimuli were heard as speech or nonspeech. Accordingly, Professor Dorman argued for the operation of two modes of processing nonspeech signals having speech-like properties. Furthermore, Professor Dorman implied that the dissociation of these two modes could be assessed by looking at differences in the location of category boundaries when the same stimuli are perceived as speech or nonspeech.

Professor Massaro departed from his symposium contribution by focusing on his general model of auditory information processing which postulates both structures for storage of information in memory and processes for carrying out various operations on this information. According to Professor Massaro's model, the earliest stage of processing involves acoustic feature analysis and is similar for speech and nonspeech signals alike. Processing here is not influenced by higher-order knowledge or context from long-term memory. Professor Massaro claimed that his general model could account for the differences observed in perception between speech and nonspeech without assuming the existence or operation of a specialized "speech mode" of processing. According to Professor Massaro, a listener's higher-order knowledge and his experience with speech affects the way acoustic features are treated and integrated at what he calls the primary stage of recognition in his model. Thus, a two stage model is also assumed to be necessary for perception of speech stimuli although the same two processes may be employed with other nonspeech stimuli.

Professor Liberman's remarks on duplex perception were summarized very briefly by Professor Studdert-Kennedy.<sup>1</sup> Using a variation of the so-called "Rand Effect", Professor Liberman has shown that listeners can simultaneously perceive a phonetic event (i.e., a CV syllable) and an auditory event (i.e., a chirp). Professor Liberman has argued that these results imply that both auditory and phonetic processes are carried out together simultaneously in parallel and that a distinct phonetic subsystem exists for processing speech signals, a subsystem which is separate from processes used to perceive other auditory signals.

Dr. Summerfield summarized his symposium paper with Dr. Bailey by emphasizing that the information for phonetic perception must be found in the acoustic signal itself which reflects the consequences of articulation of speech. Dr. Summerfield suggested that the phonetic information in the signal could be properly characterized by detailed examination of the articulatory control that gives rise to acoustic patterning in speech production and by a detailed examination of how the distinctiveness of this articulatory patterning is enhanced by auditory processing of speech signals.

-----  
1) Professor Liberman was not present at the congress.

Dr. Summerfield emphasized that this research strategy would be possible without having to assume any need for articulatory mediation in speech perception.

#### DISCUSSION

Following the individual summary statements, there was a general discussion among the panel members which was then opened up to the audience for additional questions and comments. Several broad and narrow issues appeared to emerge from the symposium papers and summary presentations as well as from the preliminary discussions that the panel members held before the symposium began.

Professor Studdert-Kennedy summarized these issues briefly before beginning the panel discussion. The first, and perhaps most general issue, concerned comparisons made in perception between speech and nonspeech signals. Specifically, it appeared that everyone agreed more or less that speech perception is in some sense special although not everyone agreed on precisely in what way it is special. Thus, the question of whether speech is a special process is one that still remains and apparently is one that continues to occupy the attention of numerous investigators working in speech perception even today.

Closely associated with the speech-is-special issue is a set of somewhat more narrowly defined experimental issues related to how one would be able to demonstrate clearly what the presumed special properties of speech are. That is, some concern was expressed among several members of the panel with the currently available methods and research paradigms used in speech perception research, particularly the use of discrimination procedures to assess differences between speech and nonspeech signals. During Dr. Ades' summary statement and later during the panel discussion, he repeated his dissatisfaction and skepticism with the traditional methods of comparing identification and discrimination of speech and nonspeech and consonants and vowels.

Another, somewhat broader issue that emerged from these discussions concerned the question raised by Summerfield and Bailey in their paper of whether there are, in fact, "characteristic" acoustic properties of speech signals that result directly from articulation and whether these properties are distinct from the properties of nonspeech signals. This particular issue highlights

the clear separation of views that emerged at the symposium by Divenyi and Massaro, for example, who suppose instead that there really are no distinctively different or unique acoustic correlates of speech sounds that separate them from the class of nonspeech signals in the listener's environment. According to both of these investigators, differential processing by a human observer is not required or determined by properties of the signal itself but rather by experience, training, context and higher-order knowledge. The early stages of perceptual processing are therefore the same for speech and nonspeech signals alike.

Finally, the issues surrounding the development of speech perception, particularly the recent findings with young prelinguistic infants, were also cited as a potentially important topic for further discussion. Professor Studdert-Kennedy wondered to what extent it is reasonable to suppose that an organism such as a young infant who does not "know" a language can respond to an acoustic signal as though it were conveying language--that is as though the signal were speech.

The panel discussion began with several additional remarks about the use of discrimination paradigms in speech perception research. Dr. Ades suggested that he could see little use for additional discrimination experiments in the future. Dr. Divenyi repeated several of his earlier comments on the need for two stages of processing in auditory perception to deal with all the relevant empirical phenomena in the literature. Moreover, he restated his claims again about the role of perceptual strategies in determining what a listener focuses his attention on in speech perception.

In responding to Dr. Ades' remarks about discrimination testing, Dr. Massaro felt that discrimination experiments should proceed in parallel with categorization experiments to illuminate the nature of processing speech and nonspeech. Moreover, Dr. Massaro summarized the results of recent experiments that manipulated several acoustic cues at the same time in order to explore how listeners integrate or combine information in complex multi-dimensional signals.

Professor Studdert-Kennedy suggested that the discussion seemed to point toward general agreement about the need for levels and stages of processing in perception, particularly speech perception. Professor Studdert-Kennedy also noted at this time that

one of the major reasons for postulating two levels in speech perception was the earlier work of Fujisaki suggesting the possibility that two kinds of auditory memory or coding were operating in categorical perception experiments.

The discussion then turned to the issue of how speech is distinguished acoustically from nonspeech signals. Dr. Summerfield pointed out that the contrast between speech and nonspeech might be more profitably examined in terms of different styles of processing--one appropriate for real world "events" (i.e., speech signals generated by a human vocal tract) and the other being appropriate for a relatively unnatural mode of processing where the object of interest is a "nonevent". Dr. Summerfield also suggested that there are reliable acoustic markers in the speech signal that inform a listener that the signal is speech rather than nonspeech. For example, the posture of the vocal apparatus during speech production is unique to speaking. There are both short- and long-term changes in variations in intensity and rise-time which are indicators of speech that may act as "trigger-features" to engage a speech mode of processing.

Professor Dorman then suggested a possible experimental paradigm to compare speech and nonspeech more directly by examination of "trading relations" between different types of acoustic cues in both contexts. If the trading relations differ between the two contexts, speech and nonspeech, then one could argue for distinctly different modes of processing for speech vs. nonspeech signals.

After the members of the symposium panel completed their discussion of these issues, the moderator opened the discussion to members of the general audience in attendance. Professor Stevens raised the issue again of what markers or characteristics distinguish speech from nonspeech signals. Professor Stevens suggested that it is not necessary to make reference to articulation in speech perception because all speech signals have three or four criterial acoustic properties that set them apart from all nonspeech signals. The first property involves the rate of amplitude variations over time. A basic property of speech is that it has a syllabic structure creating amplitude fluctuations between consonants and vowels. A second property of speech is shown in the spectra of speech signals. If the spectra of speech are sampled

at any point in time, the resulting analysis will display characteristic peaks and valleys. A third property of speech is the fact that these spectra change with time. That is, there are well-defined acoustic correlates to the changing articulatory gestures in speech production. The spectra of speech can also change rapidly or slowly over time. Professor Stevens suggested that one might speculate that speech signals are acoustic signals that the auditory system "likes" because it is easy to extract properties from signals of this kind.

Dr. Waterson then raised the question of the usefulness of the present kinds of experiments carried out on speech vs. non-speech. She argued that almost all of the research has used European-based languages with either European or American subjects and the tests employ language-specific features such as VOT. That is, the contrasts are presented in the language of the subjects. She wondered what sorts of results would be obtained if the subjects were presented with sounds from more exotic languages.

Professor Kuhl questioned the claim made earlier in the introduction by Professor Pisoni concerning the chinchilla's apparent inability to discriminate some of the cues to place of articulation in stop consonants. Professor Kuhl pointed out that the chinchilla's failure to discriminate /d/ from /g/ is due to a basic sensory limitation involving the length of their basilar membrane and not any inherent perceptual or cognitive limitation. Professor Kuhl also took issue with another remark of Professor Pisoni's in his introduction concerning the usefulness of certain kinds of comparative designs involving animal subjects and what these results could provide for understanding human language. Professor Kuhl stated that very pertinent information about "processing" species-specific acoustic signals may be provided by looking at animal models, particularly animals in which "vocal learning" is a salient characteristic such as the acquisition of bird song or coos by certain species of monkeys. Unfortunately, Professor Kuhl did not provide any further details about precisely what kinds of information would be obtained from these animal studies nor how the perceptual processing by these animals could be compared to the analyses carried out by humans.

Professor Kuhl also touched on the issue of a predisposition for processing certain salient acoustic attributes by human infants.

Such salient properties might serve to "focus" the infants' attention on certain aspects of the speech signal at a very early age. Moreover, Professor Kuhl repeated the suggestion, made by several others, that there is the strong possibility that the selection of speech sounds in language was guided, in some sense, by evolutionary constraints on the close match between both speech production and speech perception.

Dr. Klatt pointed out an important methodological difference in the results presented in the introduction by Professor Pisoni and the findings obtained by Professor Dorman on sine-wave analogs of CV syllables. Professor Pisoni showed well-defined labeling data for three categories of FMs corresponding to rising, level and falling, whereas Professor Dorman only reported two categories corresponding to rising and falling. Dr. Klatt suggested that this is a potentially important issue worthy of further study with fine-grained discrimination techniques which reduce the use of category labels. Dr. Klatt raised the question again of whether speech signals are somehow structured along "natural" auditory or psychophysical distinctions and/or constraints from the way speech is produced by the articulatory system.

Professor Fourcin offered an additional property, variations in fundamental frequency, that should be added to Professor Stevens' list for distinguishing between speech and nonspeech signals. Professor Fourcin also emphasized the need to look at pattern learning as the abstraction of invariants in complex stimuli, a topic that received little, if any, attention by members of the symposium.

Following the questions and comments from the audience, each of the panel members provided several additional final remarks elaborating on the statements they made earlier or commenting on some specific item raised in the general discussion. For the most part, however, the symposium on speech vs. nonspeech served to solidify a general sense of agreement among various investigators as to the value of comparisons in perception between speech and nonspeech signals. The issue of whether speech is special was discussed extensively throughout the symposium and led to a consensus that such a broad distinction is no longer meaningful, although nearly everyone believed that speech perception was somehow special or unique in its own way. A central issue that emerged

from this symposium was a concern with identifying the distinctive acoustic properties of speech signals that set them apart from other nonspeech signals in the listener's environment. There was also some attention devoted to questions of perceptual development in infants and issues surrounding perceptual predispositions for processing speech signals. Finally, there was a continued lively debate and interaction throughout the symposium on research methodology, particularly the use of discrimination paradigms in speech perception and the relevance of these sorts of data to categorization and recognition of phonemes in speech.



## WORKING GROUP: THE SYLLABLE IN PHONOLOGICAL THEORY

Organizer: Alan Bell

## ALAN BELL's SUMMARY

The Working Group met twice during the Congress to discuss selected issues related to the controversial unit of phonetics and phonology. The discussions largely concerned questions raised by the following papers, which the authors had exchanged among themselves and a few other researchers before the congress.<sup>1</sup>

- Árnason, Kristján: "A diachronic look at the syllable"  
 Bell, Alan: "The syllable as a constituent versus organizational unit"  
 Bell, Alan: "The role of segment bonds in phonological organization"  
 Brend, Ruth: "The syllable in tagmemic analysis"  
 Coates, Richard: "A point of universal phonotactics?"  
 Coates, Richard: "The categories of real phonology in relation to the syllable"  
 Coates, Richard: "Some allegro syllabic consonant processes in English"  
 Coates, Richard: "Reservations on the origin of syllabic consonants"  
 Cochran, Anne M.: "Notes on current research on the syllable in Papua New Guinea languages"  
 Cochran, Anne M.: "Ampeeli-Wojokeso consonant clusters--a study in syllable complexity" (with Edith and Dorothy West)  
 Galton, Herbert: "Interrelations between the open syllable and the phonological system as illustrated in Slavic"  
 Mikuš, Radivoj: "Vers une nouvelle phonétique"<sup>2</sup>  
 Price, Patti Jo: "What is the syllable anyway?"

The workshop was also fortunate to have the participation of the following Congress attendees with research experience on the syllable and related matters: H. Andersen, B. Andrésen, C.-J.N. Bailey, R. Bannert, H. Basbøll, R.A.W. Bladon, J. Bybee Hooper, W. Dressler, O. Fujimura, J. Gvozdanović, J.T. Jensen, C.-W. Kim, I. Lehiste, B. Lindblom, L. Menn, L. Papademetre, E. Pike, L. Selkirk, E. Strangert, S. Vater and K. Williamson.

- 
- 1) Requests for copies of papers should be addressed to the individual authors.  
 2) R. Mikuš was unfortunately not able to attend the Congress.



The first session opened with discussion of Price's experiments on the acoustic cues sufficient to shift identification of tokens prepared with the aid of speech synthesis among prayed-parade-braid-bereted and among plight-polite-blight-belight. This led to a general discussion of a wide variety of such phenomena, including some of particular interest mentioned in Cochran's paper, and of acoustic cues involved. Some comment on the different ways judgements on the number of syllables can be obtained also followed. Discussion then turned to the concept of the relative "resistance to coarticulation" of segment classes presented by Lindblom and Bladon and to a theory of the internal structure of the syllable sketched by Basbøll. The session concluded with discussion of Coates' proposal that the syllable functions as a domain of feature timing in a phonological theory in which time rather than sequence is the basis of phonological representation.

The first topic of the second session was the role of the syllable in diachronic phonology, under which three cases were taken up. These were Galton's contention that the open syllable canon of Slavic was a principal factor in the development of the correlation of palatalization, Árnason's study of vowel shortening and lengthening in Icelandic, which he concluded to be inadequately explained by several different theories of syllabic representation, and the case of cluster formation in Modern Greek presented by Papademetre. The final topics of the workshop were Bell's proposed framework of segment bonding as an alternative to current syllabic models and the general question of the hierarchical nature of the syllable as described in tagmemic theory by Brend.