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 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

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

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 suprasegmentals, 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 Properties

a. 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 <u>inspiration</u> when lung volume is smaller than its resting level. For many years now the work of Ladefoged and his 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 Bernouilli 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 coordinate 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 o: 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 midsaggital 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 intraoral 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. Nooteboom, 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 <u>and</u> 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 [ɹ] 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. "tointed 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 control 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 /ɛ/ 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 consider 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, B. (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. Öhman (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 Wicksell. 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

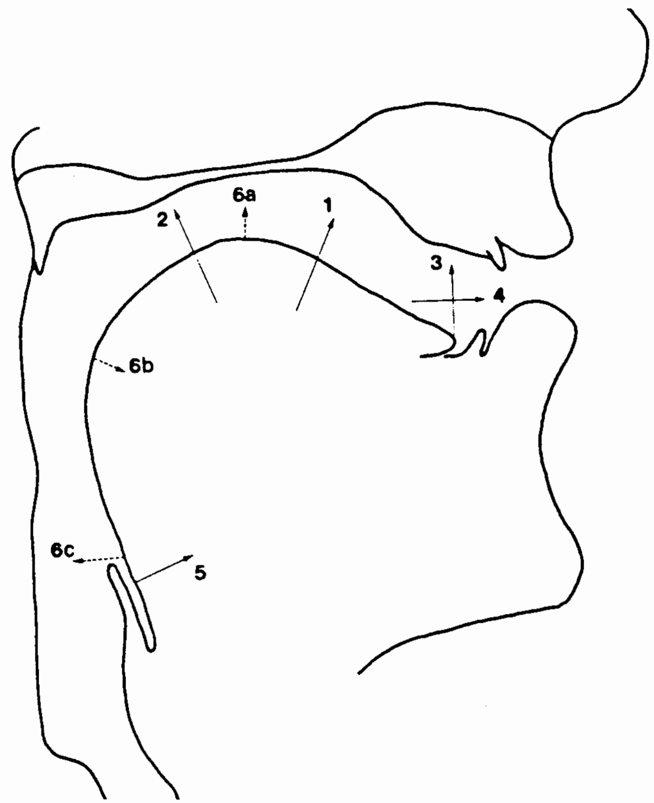


Figure 1

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 /əɹ/ 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. Öhman (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. Öhman (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 actitivy, 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.

Benguerel, A-P., H. Hirose, M. Sawashima and T. Ushijima (1978): "Laryngeal control in French stop production: a fiberscopic, 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 Phoniatrics, 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.

Speech Perception

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 adjutant 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 Nooteboom 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; Nooteboom, 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: Nooteboom 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, Nooteboom 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'Shaugnessy 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, Zhukov, 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 ɛ a ʌ 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 construction. 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 aftereffects 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 [splɪt] (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, Rudnicky, 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 /bɛd/, /bæd/, /bɛt/, /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. Nooteboom (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 Nooteboom, 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 Nooteboom, 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 Nooteboom, 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 vospriiatiie, 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 psychoacoustics. 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 (F1 - steady-state, F2 - time-varying), subjects were able to locate the point corresponding to the shift from [iru] to [içu]. 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_2. Subjects could do this just as easily when the time-varying F_2 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 A1/A2 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) F2 contours revealed three cues used in phonetic interpretation (Lublinskaja and Slepokurova, 1977; 1978). One cue corresponds to F_2 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_2 or of the spectrum shape. The third cue is the direction of the initial F2-transition. This last cue appeared to be effective only in a restricted frequency range since it only serves to differentiate between [ʉ] and [ɨ] and between [ø] or [œ] and [ɛ] or [ə]. The boundary in the direction of the F2-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. Öhman (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 admittable 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) (Kuwahara 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 (Nooteboom 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. Öhman (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. Nooteboom (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. Nooteboom (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.

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.

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

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

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.[1] 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 phonology, 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[2]) a "school" in this sense[3], 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.[4]

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"[5], 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, 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.[6] 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-

---

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).

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 specifiable 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).[7]

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.[8] 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)[9]: 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[10] 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.[11] 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 glossemes. 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).[12] 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.[13]

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).[14] 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 Rudes 1976, where the name 'cycle' is coupled to stylistic variation), which has been argued for by Brame 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 1972.[15]

In several papers (e.g. Basbøll 1975 with reference) I have generalized 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 delineated by the boundary in question (or a stronger one) as its domain, and that no boundary should be allowed to occur properly included in the SD of a rule. Furthermore, rules of a lower rank seem to apply quite generally before rules of a higher rank, and this may be taken as a suggested further principle of precedence if one works within a theory of "no extrinsic order".

3. Metatheory and evidence in phonology

In his suggestive "Epilogue: Linguistics as metaphysics" with the subtitle "on the rationality of non-empirical theories", Lass (1976, 213-221) uses Popper's well known demarcation criterion for "scientific" (as opposed to "metaphysical" or "philosophical") theories: that they must be falsifiable (or refutable). Lass concludes 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 <u>rationally arguable</u>: 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 <u>prognostic</u>, 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-

15 Rischel proposed that the branches of an ordinary tree structure were either plus (= strong) or minus (= weak), and he showed how some very simple principles permitted the actual stress pattern of e.g. Danish compounds to be read directly from the tree structure (without any erasure of structure followed by the re-application of rules). What is essentially the same solution has been reproposed by Liberman 1975 (apparently without any knowledge of Rischel's paper, but cf. now Liberman and Prince 1977).

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 sociolinguistic 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[16] 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"[17] (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.[18]

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 recommendable 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 botn "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[19] 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 Praguian "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.

Davidsen-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.[20] 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.[21] 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 fonctionelle, 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.: Anma 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. (1955): 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.

Liberman, M. (1975): The Intonational System of English, Bloomington, Indiana: Indiana University Linguistics Club.

Liberman, 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'archiphonème 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 'logicist' 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 /kəmba/ 'insipid', the elative is kəmbu. 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 /kəmba/ 'insipid', the elative is kəmbu. 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 *Lautgesetz* 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 elative formation", in Studies in the linguistic sciences 4:2, 91-111.

Gussmann, Edmund (1978): Explorations in abstract phonology, Lublin: Uniwersytet Marii Curie-Sklodowskiej.

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. <u>The "abstract" positions</u>

2.1. The tradition of the <u>Sound Pattern of English</u> (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, <u>such a supposition seems to be based upon an excessive reliance on surface data</u>" (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 <u>SPE</u> 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 <u>SPE</u>. 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 <u>SPE</u> 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 morphological 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→B, no language will contain the converse rule B→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.

Davidsen-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. Juilland (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.

Special Lectures

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-

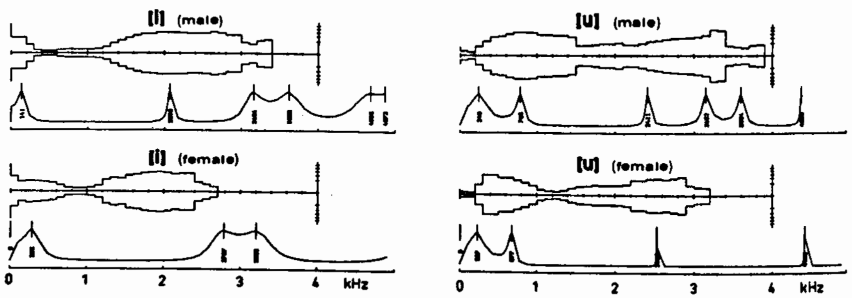


Figure 1

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 [ɩ]. 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 Δx 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 2 m 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 quantizied in 2 m sections of equal length and there could exist a unique solution for the 2 m 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 cm^2 and a length 8.7 cm, and the front tube an area of 1 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, F1 F2 and F3 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 automatic 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 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. I, I-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 x B.
- La frontière phonémique n'ést 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).

Réference

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'électropalatographes, 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 électropalatographie 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,[1] L. J. Raphael,[2] D. B. Pisoni,[3] and J. R. Sawusch,[4]
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,ɩ,e/ or /ɛ/, 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-ɩ/ and /e-ɛ/ 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, Université d'Amsterdam, Singel 134, 1015 AG Amsterdam, Pays-Bas.

A partir d'un corpus d'arabe classique, nous avons étudié les manifestations de l'emphase et de la gémination. L'entourage des occlusives /t tt ṭ ṭṭ/ se limite à la voyelle /a/. Nos résultats proviennent de 2 radiofilms (42 images/sec), d'oscillogrammes à 4 lignes et de sonagrammes en filtrage large. Les films ont été mesurés grâce à la méthode des axes, tracés de 20 en 20 degrés.

Résultats et discussion

Le rapport des durées géminée/simple est de 1.76 pour les emphatiques et de 2.02 pour les non-emphatiques. Les géminées présentent un renforcement de l'articulation et une accentuation des propriétés des consonnes simples: pour /tt/, fermeture et antériorisation; pour /ṭṭ/, augmentation du caractère emphatique (extension de la constriction postérieure en direction de la zone uvulaire). On ne relève aucun mouvement spécifique de la paroi postérieure du pharynx, ni pour la simple - nous sommes en accord avec Ali et Daniloff (1972) - ni pour la géminée. L'indice de dispersion articulatoire (langue et maxillaire) est moins important pour le degré long de quantité, qu'il s'agisse des emphatiques ou des non emphatiques. Quant aux voyelles, leur articulation diffère selon qu'elles sont au contact d'une simple ou d'une géminée, et de ce fait, leur structure acoustique est soumise à quelques variations.

Enfin, les données dont nous disposons, et particulièrement l'examen image par image des radiofilms, ne montrent aucun signe de réarticulation (Lehiste, 1973) de la géminée. Non seulement on ne constate aucune marque de fléchissement durant la tenue, mais de plus la stabilité organique est très supérieure. Néanmoins, les configurations articulatoires différentes et la faible dispersion des mesures semblent révéler une programmation spécifique des contrôles moteurs pour les géminées et leur entourage immédiat.

Références

Ali, L.H. et R.G. Daniloff (1972): "A cinefluorographic investigation 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˪/, /ʃi/, and /zi/, while others were less familiar /py/, /pø/, /χi/, and /ɣi/. 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_o/T, where T_o 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 incisives (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 Geršić, 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 Fernsehbildschirm 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 [Φ] 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 [Φ], 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_o. 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_2O 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 misreproductions 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 Phoniatrics, 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. Condax, 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 fiberscope 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[1] 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 interstimulus 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 FRANCAIS 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'électropalatographie. Notre système composé d'un palais à 64 électrodes et directement relié à un mini-ordinateur sera présenté lors de l'exposé 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 articulatoire des consonnes. Les conséquences acoustiques des mouvements de la langue ont été interprétées à partir d'une analyse sonagraphique.

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 significativement 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'énergie dans le spectre alors que l'occlusion articulatoire n'est pas complète. 4) les occlusives sourdes possèdent une longue phase implosive 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 position 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

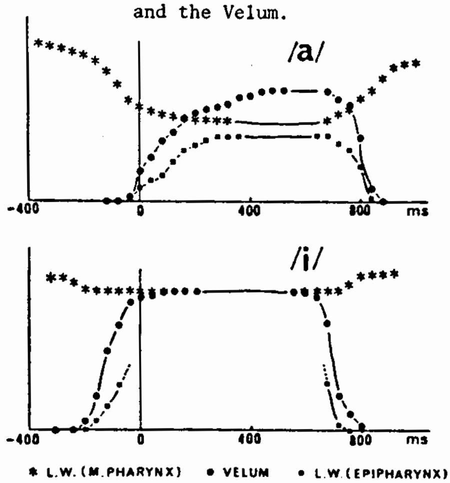
Seiji Niimi, Research Institute of Logopedics and Phoniatrics, University of Tokyo, Japan

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).

Displacement of the Pharyngeal Walls and the Velum.



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.)

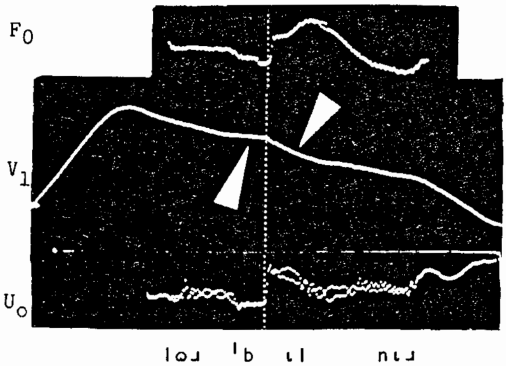


Figure 1. Averaged fundamental frequency (F_o), lung volume (V_1), and oral airflow (U_o) during the utterance 'Lure Bill near', [lɑɹ ˈbɪl nɪɹ] as produced by an adult male speaker of English. The dotted vertical line marks the synchronization point used to form the averages.

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 occulusives 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 /ɲ/ 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 aperture, 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 aperture, 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[1], 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 ± 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 sumultaneously 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 sounas 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 RADIOCINEMATOGRAPHIQUES 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^{h}yu/ →[gaču]), 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^y/. 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 /sš/, 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, ɛ, a, ɔ, u, ɜ/ 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 Andrésen, 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-Ḥâmma, 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, halflong 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, Rijksuniversiteit 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 [l̩], [n̩] and [m̩]. These signals were processed by a PDP-12 computer and the relevant portions of the signal were isolated. The durations of the [ɝ], [u], [l̩], [n̩] and [m̩] 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.

Carpeneter, 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 realiably 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 [ṋ]); it contrasts with a nasalized stop [n] 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.[1]

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 Connecitcut,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 [ɑ].

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 [ɑ].

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 [a] and M.E.H. Schouten, Institute for Perception TNO, Soesterberg, the Netherlands. [a] 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 dienten 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, sodass 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 $/\text{ba}/_R$ vs. $/\text{ga}/_L$. 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 avions 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 proposions 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 imposee 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

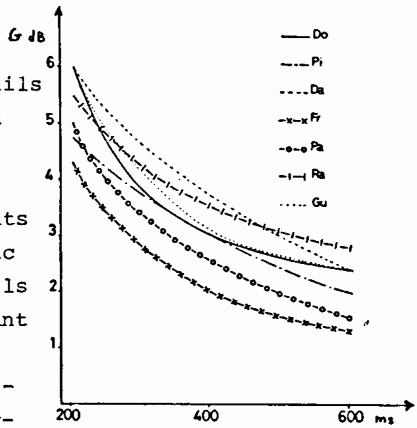
Christel Sorin, CNET, Lannion, France

Il est généralement admis que l'étude des faits prosodiques doit se faire au niveau des trois paramètres: fréquence fondamentale (F_o), 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 interpretation



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, Phoniatrische 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.

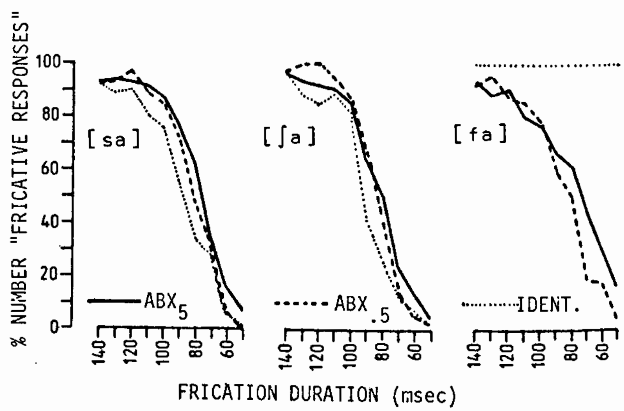
Ulbrich, H. (1972): Instrumentalphonetisch-auditive R-Untersuchungen im Deutschen, Berlin: Akademie.

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. Öhman (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.

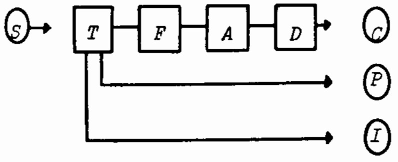
Nagai, K. et al. (1956): "Analysis of phonemes by articulation tests", Journal of the Acoustical Society of Japan 12, 148-154.

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 fundaments 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): Nasalitätsdetektion und Vokalerkennung (=Forum Phoneticum 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 2π 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, ɥ, 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,ɥ,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,ɥ,w/ n'est sujette qu'à des variations minimes 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,ɥ,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 Fø 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,ɥ,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 Lancia, 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_o 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_s 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_o. Ensuite, le couplage a été simulé pour différentes valeurs de P_s (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 ɛ œ a ɑ ɔ o u]. Les résultats montrent que pour F_o voisin de 120 Hz, le couplage introduit des variations de l'ordre de 8 Hz; F_o 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 ɛi, ai, ȍɪ, ɔu and ʌü and 6 opening diphthongs iə, üə, ɩə, ʌə, uə, ȍə. 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 F1 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 F1. 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 TRANSSEXUELS: 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_2O
/u/	.54 cm H_2O
/ɔ/	.51 cm H_2O
/æ/	.43 cm H_2O

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 [ɑ]-[ɑ̃], [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 controverses 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 hierarchisé 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 (eds.): 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.", 9es Journées d'étude sur la parole, Lannion, 31 mai - 2 juin, 321-342.

ETAGE PHONETIQUE D'UN SYSTEME 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 calculateur 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 Bernouilli 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 / Informant / Word	Tonyu	Tokyo	Owari
	MT, HY	SU	SM
/he⌝bi/	-	70	100
/he⌝mbi/	-	-	170
/he⌝qbi/	200	-	-

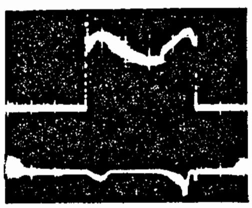


Fig. 1. The pitch contour of /he⌝qbi/ spoken by TM.

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. Aĺmukhamedova, Faculty of History and Philology, Kazań State University, Kazań, 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 [saˈbaka], is phonologically /s|o-a|báka/. 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 [ˈvodu] 'water', nom. vodá, /v|o-a|dá/, pronounced [vɑˈdɑ]. 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 Laríssa). Very often such neo-etymologizations appear in jargon (like lor for laringólog), in children's speech (like [ˈflaki], plur. of flag /fla|g-k|/, pronounced [flak] 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 [ˈvototʃkɑ] in dialectal speech (in the orthography vodočka) from [votka] /vó|d-t|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 [± length] , or implicitly, e.g. in such features as [± 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",

$\underline{\underline{da}}:\underline{\underline{ri}}$ "he used to",

$\overline{\overline{da}}:\underline{\underline{ri}}$ "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.[1] 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 Trubetzkoyschen Regel müssten auch ě und ó 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 ó, 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. Trubetzkoyschen 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, wɨw, ɥɨh/. 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 (111 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 "Pfalzsches 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 hihhuli 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 (geminates) 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 suprasegmental. 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 paradigma (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 [bogə̑š], teils durch Assimilation zu den betonten Vokalen [o] und [u] [norŏt] 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 [sožéŋ]. 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 [ɯ]; 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 > ɨ] et [u > y]. [y < u] aboutit souvent rapidement à [ɨ] (grec ancien, etc...), ce qui suggère plutôt [u > ɨ], 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^e 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 ç, 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 ç 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-murmered 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-Hůrková, Czechoslovac 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 jezyka 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^h] and [b^h], 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^h] are produced with slack vocal cords, fundamental frequency (F_o) on the vowels flanking medial stops and sonorants was sampled and averaged over 10 tokens of each consonant type. F_o was significantly lower on the vowel following [b^h], but there was no appreciable perturbation of the F_o 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^h] 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 Schreibungssystems 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") /š ŋ pl/. 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 Ducth 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 /ʃ,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; /œ:/ and /y:/ 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

$$(1)\quad V \rightarrow [\alpha\ \text{back}]\ /\ \begin{bmatrix} V \\ \alpha\ \text{back} \end{bmatrix} C_o\ \_\_$$

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 spekers 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[1] 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 Explosiongeräusches - 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,[2] wo auch die Hierarchie der Eigenschaften ausgedrückt wird, kann man die Struktur dieses Merkmals folgendermassen darstellen:

$$\left\{ \begin{array}{c} TF_2 \\ CEx - QEx - TF_1 \\ (QCons) - (TF_3) - \ldots \end{array} \right\}$$

---

(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 Meinhert 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 triphonemic 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 nth 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: $\frac{a:b}{Pr} \leftrightarrow \frac{C}{Pn}$ (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 FRANCOPROVENÇAL

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 [ɫ] 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 [l̥].

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.

/ṭs/: 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 /ṭs/. This paper reports the existence of /ṭs/ 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^h/, /tʃ/ and /dʒ/ which represent aspirated, unaspirated and voiced palato-alveolar affricates, respectively.

It is worth mentioning that /ṭs/ 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 /ṭs/, F_1 and F_2 behave exactly in the same manner as with other well-established emphatics both in Neo-Aramaic and Arabic, in that F_1 is raised while F_2 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 /ṭs/ when compared with the German /ts/, and partly on a comparison with the Aramaic /tʃ/ and /tʃ^h/[1] (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 /ṭs/ is nearer in magnitude to that of /tʃ/ than to that of /tʃ^h/. This phonetic similarity tempts one to envisage that /ṭs/ 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^h/, 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; b̰, a voiced, zero pressure implosive; b, a voiced, non-glottalized variant; p<, a voiceless implosive. Alveolar implosive variants are: ɗ, 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 ṛ IN PIGNASCO

Gladys E. Saunders, University of Virginia, Department of French and General Linguistics, Charlottesville, Virginia 22903, USA

The nature of retroflex ṛ and its interaction with non-retroflex resonants is examined in Pignasco (a Gallo-Italian dialect spoken in western Liguria). It is shown that retroflex ṛ 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. [boːṛa]-[boŋ] 'good', fem. sing. and masc. sing., respectively). It is argued that the retroflex ṛ 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 ṛ 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 disjective) 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 → ʔ / __ ## (loss of oral contact for /t/)
2) t → Ø / __ s## (final cluster simplification)
3) t ## j → tʃ (affrication)
4) ð → Ø / apical continuant # ___ (ð 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/, /č/, /q/; /p^h/, /t^h/, /k^h/, /$č^h$/, /q^h/; /$p^?$/, /$t^?$/, /$k^?$/, /$č^?$/, /$q^?$/. 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^h/ vis-a-vis /$p^?$/, /t^h/ and /$t^?$/, etc..

We found that for all pairs the development of $C^?$ implied the development of C^h, and that C^h 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 [ɹ] or the rolled alveolar [r]; and the retroflex L2 [ɽ] or the frictionless [ʁ] 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 [r ɽ ʁ], and [ɽ ɹ] 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. Denes, 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_o) pattern was obtained from the glottal excitation amplitude (A_o) pattern by means of the following rule:

$$F_o = FO \frac{A_o + a}{1 + a}$$

where FO is the F_o target value (Hz), and a is a numerical coefficient. A phrase intonation contour, F_{oc}, was obtained by multiplying F_o 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. Blasdell (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, ɟ, 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 [''ɑʈa:mɑʈ 'ɖʝa: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(pitch), log(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 SYNTHESE AUTOMATIQUE PAR DIPHONEME

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 marque[illegible]s prosodiques. Le traitement des homographes qui ne peut être fai[illegible]e 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 dèterminent 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 Wroclaw, Wroclaw, 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, \ldots 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[1]

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 rhythmicality (i.e. with greater emphasis on rhythm).

The results are as follows:

1. Emphasized rhythmicality 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 tout 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. Anôther 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.

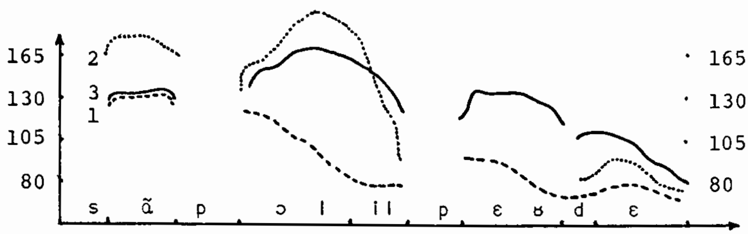
A non-phonematic glottal stop ([ʔ])in the Russian language and in a number of its dialects marks the initial stressed vowel. In this way, it underlines the beginning of a word.

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 ambigües, 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_o) 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_o values within each syllable, F_o at the starting point, F_o at the end point, F_o at the last turning point, maximal F_o 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 $\theta = \langle P_1, \ldots P_T \rangle$, where T represents the number of initial parameters, $P_1, \ldots P_T$ are weight coefficients. If vector θ differentiates the opposed pitch contours it can be considered as a distinctive feature, while each $|Pi|$ $(1 \leqslant i \leqslant T)$ value can be viewed as estimates of the initial parameters' significance. Vector θ was computer determined in a manner whereby all the pitch curves' values of statements in reference to θ 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_o 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 α, ω and ϕ 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$\overset{2}{\text{e}}$nner-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-l$\overset{1}{\text{i}}$ten* ('too small'). In isolation, *liten* has tone 2, but assuming that *for liten* 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$\overset{2}{\text{a}}$ta*, *h$\overset{2}{\text{y}}$tta*, whose *indefinite* forms are *g$\overset{2}{\text{a}}$te* ('street') and *h$\overset{2}{\text{y}}$tte* ('cabin'), both bisyllabic, as opposed to the monosyllabic indefinite forms in the cases where we get tone 1. A simple segmentation of *g$\overset{2}{\text{a}}$ta* 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'ambigüité 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 Fo: 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 (Fo_i) of vowels in various different languages. These studies consistently indicate a strong inverse correlation between Fo_i and the first formant of the vowel. The coefficient of determination (R^2) between F1 and Fo_i for the data given by Peterson and Barney (1952) is 0.85. Calculating the regression line from F1 to Fo_i consequently gives a reasonably close estimation of Fo_i. This estimation can be considerably improved if we take into account the second formant (F2), since we obtain an R^2 of 0.922. An even better correlation is found between Fo_i on the one hand and F1, F2 and $\overline{Fo}$ (the mean Fo for each subejct) 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 $Fo_i = a_o + a_1\overline{Fo} + a_2F1 + a_3F2$ where $a_o = 20.166$, $a_1 = 0.975$, $a_2 = - 0.034$, $a_3 = - 0.002$ 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 rhytmischen 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 rhytmischer 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 rhytmischen 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 rhytmische 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): "Rhytmic 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 sp$\overset{1}{e}$g$\overset{2}{e}$l 'mirror', which also have accent II. The THEME stress rule operates in words like gor$\overset{1}{i}$ll$\overset{2}{a}$ '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 j$\overset{1}{o}$rdand$\overset{2}{e}$ 'burying' and j$\overset{1}{o}$rd$\overset{2}{a}$nde 'earth spirit'. The COMPOUND rule, properly formalized, predicts accent I for compound verbs of the type bet$\overset{1}{a}$la 'pay', although the basis verb t$\overset{1}{a}$l$\overset{2}{a}$ 'speak' has accent II. This analysis of betala can be generalized to adjective phrases of the type för många '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 (Trubetzkoj, 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

Trubetzkoj, 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.

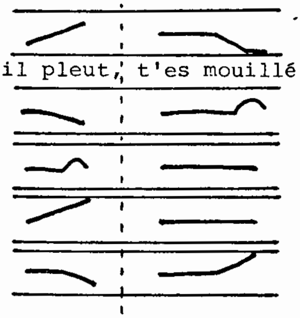
ASYNDETE ET INTONATION EN FRANCAIS.

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 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 Paryže (But he wasn't in Paris, you are wrong);
Tol'ko on ne byl v Paryže (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 coexist 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 Fo, 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 anacruses themselves and of rhythmical units with anacruses, 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 anacruses 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 anacruses. A great prevalence of conjunctions is to be found in the function of anacruses.

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 KV́K und KV́: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 KV́K und KV́: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'élision 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 reinterpretation 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 russkogo jazyka, Moskva: Izd. MGU.

Bryzgunova, E.A. (1977): Zvuki i intonacija russkoj reči, 3-e izd., Moskva: Russkij jazyk.

Romportl, M. (1971): "K synonymii a homonymii intonačních prostředků", 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 Uberprüfung 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 melodical 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.[1]

Conclusion

A preliminary survey of the results leads us to the conclusion that the stylistic heterogeneity of the texts determines their peculiar accentual and melodical 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_o movement at the beginning and, in particular, at the end of the phrase. Thus, for phrase initial and phrase final phonetic words, F_o 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_o.

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*, - F. Degan* - C. Minnaja* - L.G. Paccagnella*, *Centro di Studio per le Ricerche di Fonetica (C.N.R.), Padova, Italy
*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 lenght 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 /ɑ/ 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 articulatoire, 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 glottalisation, 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 phoniatrics, 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 phoniatric-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 phoniatrics at Huddinge University Hospital and the Speech Transmission Laboratories at the Royal Institute of Technology, Stockholm. Clinically experienced logopedists and phoniatricians 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 Kreiskrankenhauses, 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 phoniatrischen 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 lanque 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.[1] 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.[2] 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 "ɣunnah" 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, pratipādya, 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',[1] 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 Vriti, 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. Zonneweld (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 *Alfabet 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 biased 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 Anthrop., 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 PHONETIQUE 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^e 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'OUJDA (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 panchronique, 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 discontinous 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 FRANCAIS DE MONTREAL; ETUDE SOCIO-PHONETIQUE

Laurent Santerre, département de linguistique, Université de Montréal

A partir de 51 interviews d'une heure de locuteurs francophones 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 constraintes 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'échantionnage 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.[1] 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_{const}", 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 govor) allows to state that they have the following articulatory characteristics (zones, height, labialization):

Russian letters	vowel sound	IPA	articulatory characteristics
а	≪ α̇: V ≫	[ä]	back zone, very much advanced, low (the V-th grade), non-labialized.
о	≪ ɒ̇: V ≫	[ɒ+]	back zone, very much advanced, low (the V-th grade), labialized.
у	≪ ʊ̇: II ≫	[ü̞]	back zone, very much advanced, high (the II-d grade), labialized.
е, э	≪ ɛ: IV ≫	[ɛ̈]	front zone, very much retracted, mid (the IV-th grade - a little closed), non-labialized.
и	≪ I: II ≫	[i̞]	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.