

PHONETIC UNIVERSALS IN PHONOLOGICAL SYSTEMS AND THEIR EXPLANATION

Summary of Moderator's Introduction

John J. Ohala, Phonology Laboratory, Department of Linguistics,
University of California, Berkeley, California, U.S.A.

In many ways the study of phonetic and phonological universals is a relatively old endeavor in linguistics and in other ways it is relatively new. It is old in the sense that some 100 years ago when our intellectual forefathers, Ellis, Sweet, Bell, Lepsius, Passy, Jespersen, and others were struggling to develop a workable phonetic alphabet that could be used to transcribe the sounds of any language, they had, implicitly at least, to deal with the problem of whether there were phonetic universals. That they succeeded in devising such a practical universally-applicable phonetic notation, such as we use today, is a tribute to their hard work, their vast experience with languages of the world, and their scientific judgment. The phonetic alphabet and the set of descriptive terms accompanying it are not perfect, of course, but modern work on universal sound patterns would be impossible without these important tools.

It is an old interest, too, in the sense that during the past century there has been a steady, if small, flow of relatively sophisticated explanations for observed universal sound patterns, e.g., (to mention just a few) Passy 1890, Issatchenko 1937, Troubetzkoy 1939, Jakobson 1941, Hockett 1955, Martinet 1955. Characteristic of the keen insights offered during this period are the following (and here I present remarks relevant to topics of particular interest to this symposium):

Passy (1890) on obstruent devoicing: 'On remarque que ce sont les explosives qui se dévocalisent le plus souvent. Cela se conçoit, car pour produire une explosive vocalique, il faut chasser dans une chambre fermée l'air qui fait vibrer les cordes vocales; action qui nécessite un effort considérable, et ne peut pas se prolonger. Aussi les explosives doubles sont-elles particulièrement sujettes à devenir soufflées..' [161].

Chao (1936) on the patterning of voiced glottalized stops: 'A very significant circumstance about the occurrence of [ʔb, ʔd, ʔs, ʔj] is that in all the [Chinese] dialects in which they are known to exist they are always limited to labials and dentals and never exist in velars ... The reason is not

far to seek. Between the velum and the glottis, there is not much room to do any of the tricks that can be done with the larger cavity for a b or a d. As soon as there is any vibration of the vocal cords, the cavity for a g is filled and a positive pressure is created. There is therefore no space or time to make any impression of suspension [of voicing via simultaneous glottal closure, as with ?b and ?d] or of inward "explosion" as with [b, d]. The velar plosive is difficult to voice without having to do any additional tricks.'

In another sense, however, interest in universal sound patterns is rather recent or at least renewed. This has come about, I think, due to the interaction of a number of trends and events. First, there is the interest in phonological universals stimulated by the new set of research goals presented by generative phonology, namely, to look at universals of language for what they will reveal about humans' genetically-based capacity for language.

Second, there has been the sheer accumulation of reliable phonological data on a large number of languages. Works such as Guthrie's Comparative Bantu and Li's Comparative Tai (to mention just two), which synthesize large amounts of phonological data, exemplify this trend. It is because of this latter development that a project such as the Stanford Phonology Archive, constructed by Charles Ferguson and Joseph Greenberg, was possible. Third, there has been the almost explosive growth of experimental phonetics over the past 30 years or so -- especially in the development of empirically-validated mathematical models of various aspects of the speech production and perception mechanisms. In short, phonologists have realized that the study of universal sound patterns can be interesting and very important and that they now have the resources to do a better job of it than ever before.

The contributions to this symposium on phonetic universals represent very well the wide range of data, of talents, and of theoretical outlooks that are necessary in this area.

Björn Lindblom, in 'Some phonetic null hypotheses for a biological theory of language' raises the possibility that the form of language and the range over which it varies when it changes, may be determined by the biological constraints of its human users. He looks to phonetics to provide the evidence on this issue.

Kenneth N. Stevens, in 'Bases for phonetic universals in the properties of the speech production and perception systems' considers how the natural classes among speech sounds must arise due to the individual members of the classes sharing common modes of production at the neuromuscular level and/or giving rise to a common set of sensory images via the tactile, kinesthetic, or auditory channels.

Kenneth Pike, in 'Universals and phonetic hierarchy' suggests that the inability of phonologists to integrate such elusive units as the syllable or stress group into their descriptions of language may be due to their commitment to use just a single hierarchical structure. He proposes the use of parallel but interlocking hierarchies, e.g., one each for the phonological, grammatical, and referential domains.

Two papers in this symposium and one section paper deal with closely related topics on universal patterns in languages' obstruent inventories.

Thomas V. Gamkrelidze, in 'Hierarchical relations among phonetic units as phonological universals', presents a comprehensive analysis of universal co-occurrence tendencies among various features of obstruents, e.g., place of articulation, voicing, glottalization, and uses this to support a reanalysis of the Indo-European stop inventory.

André G. Haudricourt, in 'Apparition et disparition des occlusives sonores préglottalisées', presents the phonetic factors that lead to the development or loss of voiced preglottalized stops and presents extensive supporting cross-linguistic data, especially from South and Southeast Asian languages.

Sandra Pinkerton, in her section paper 'Quichean (Mayan) glottalized and non-glottalized stops: a phonetic study with implications for phonological universals', presents instrumental data on the manner of production of glottalized stops in five Mayan languages. Having found voiceless uvular implosives, she proposes a revision of Greenberg's (1970) implicational hierarchy for glottalized stops which would equate it to Gamkrelidze's claims: voicing is marked for velar obstruents, voicelessness for labial obstruents.

Robert K. Herbert, in 'Typological universals, aspiration, and post-nasal stops', points out several universal patterns characteristic of nasal + stop clusters and uses these to call into

question one reconstruction of the history of such clusters in Eastern Bantu languages.

Jean Marie Hombert, in 'Universals of vowel systems: the case of centralized vowels', presents data from speech perception tests conducted in the field with speakers of Fe?fe? (a Bantu language of Cameroon) which suggest that the universal tendency of disfavoring central vowels may have its origin in a human auditory constraint.

In my own paper, 'Universals of labial velars and de Saussure's chess analogy', I present four phonetically-based universal patterns characteristic of labial velars and use this to call into question the wisdom of structuralist phonology's pre-occupation with system-internal relations in language and their descriptions.

Conclusion

It is worth mentioning that study of phonological universals is of more than theoretical interest. If it is done well, it could yield results of great practical benefit, too, e.g., in such areas as speech therapy, second language teaching, speech recognition, and neurophysiology.

References

- Chao, Y.R. (1936): "Types of plosives in Chinese", Proc.Phon. 2, 106-110, D. Jones and D.B. Fry (eds.), Cambridge: The Univ. Press.
- Greenberg, J.H. (1970): "Some generalizations concerning glottalic consonants, especially implosives", IJAL 36, 123-145.
- Hockett, C.F. (1955): A manual of phonology, IJAL Memoir 11.
- Issatschenko, A. (1937): "A propos des voyelles nasales", Bull. Soc. Ling., Paris 1938, 267ff.
- Jakobson, R. (1941): Kindersprache, Aphasie und allgemeine Lautgesetze, Uppsala.
- Martinet, A. (1955): Economie des changements phonétiques, Berne.
- Passy, P. (1890): Etude sur les changements phonétiques, Paris: Librairie Firmin-Didot.
- Troubetzkoy, N.S. (1939): Grundzüge der Phonologie, Prague.

HIERARCHICAL RELATIONS AMONG PHONEMIC UNITS AS PHONOLOGICAL
UNIVERSALS

Thomas V. Gamkrelidze, The Oriental Institute, Academy of Sciences,
Georgian SSR, Tbilisi, USSR

After splitting the phoneme into its minimal components - distinctive features - and viewing it as a bundle of such features the question arises as to mutual compatibility of the features within the bundle and their relationship to one another on the axis of simultaneity.

It is the differing capacity of distinctive features for relating with one another into simultaneous combinations or "vertical sequences" that creates bundles of features differing in character and possessing a varying degree of "markedness", i.e., combinations of features characterized by commonness, naturalness, high degree of occurrence in the system ("unmarked") and less common, less natural combinations of features manifesting a lower degree of occurrence ("marked"), cf. Gamkrelidze (1975).

Depending on the varying capacity of distinctive features to combine with one another in a simultaneous bundle, it proves feasible to set up a gradation scale of "markedness" of simultaneous (vertical) combinations of features. Opposite extreme values on such a "scale of markedness" involve: (a) obligatory combination of the distinctive features on the axis of simultaneity, i.e., maximally "unmarked" combinations (as, e.g., combinations of features like [+syllabic, -nonsyllabic], [-syllabic, +nonsyllabic] or [discontinuous, dental], etc.) which are represented in any phonemic system being a constituent part of the phonemes entering the minimal phonemic inventory of language, and (b) noncombability, mutual incompatibility of features potentially forming maximally "marked" combinations (e.g., the features of [glottalization] and [voice] or the features [nasal] and [fricative] that are incapable of combining into simultaneous bundles).

Between such extreme values of "markedness" are arranged all kinds of possible combinations of distinctive features with varying degrees of "markedness" - with a greater or lesser approximation to the extreme values reflecting the varying capacity of distinctive features to combine with one another in forming simultaneous bundles.

Such a "scale of markedness" of combinations of distinctive features must, in principle, be characterized by a fairly high degree of universality, for it reflects the capacity - common to human language - of definite phonetic and acoustic-articulatory properties to combine more or less freely and form simultaneous articulatory complexes. Definite phonetic features, owing to their acoustic-articulatory peculiarities, combine preferably with one another on the axis of simultaneity. "Marked" bundles of features reflect - in contrast to "unmarked" bundles - a limited capacity of definite phonetic features to join in simultaneous combinations, i.e., their lesser tendency to mutual combination. Hence such bundles represent less usual or less natural combinations of features, being placed on the "scale of markedness" closer to the maximal value of "markedness".

It is but natural to expect that such bundles (and correspondingly the phonemes represented by them) will be characterized by a lesser degree of actualization in language than will features which, in view of their acoustic and articulatory properties, combine easily with each other, representing natural or usual combinations of features. The former group of bundles of features (and correspondingly the phonemes represented by them) constitutes functionally weak units in the system, being characterized by a low degree of occurrence (frequency) and distributional limitations or being entirely absent in a number of languages, forming gaps in the paradigmatic system; the latter group of bundles is more common and natural and, in this sense, "unmarked", forming functionally strong units of the system and being characterized by a greater distributional freedom and a higher degree of occurrence (frequency) - some of them with a probability of occurrence equal to one (maximally "unmarked" combinations of features). Thus definite distinctive features combine with one another in simultaneous bundles in preference to other features, the combinations of which on the axis of simultaneity form more complex units in terms of articulation and perception. Being less optimal, such combinations are of a limited occurrence in the system, forming less natural phonemic units characterized by a lower frequency of occurrence and equalling zero in certain systems (yielding phonemic gaps in the paradigmatic pattern).

The phonemic units representing stable and "unmarked" bundles

of features in any linguistic system may be characterized as "dominant" as opposed to the less common and less natural (i.e., "marked") units of the system that may be styled "recessive". Thus, any two phonemic units opposed to each other in the paradigmatic system by the hierarchical relationship of "markedness" may be characterized as "dominant" vs. "recessive", while the relationship of "markedness" itself, implying a dependence between these units, may be restyled as the relation of "paradigmatic dominance". The terms are obviously borrowed from molecular biology, known for its ample use of linguistic vocabulary in application to the genetic code (cf. Jacob, 1977). Such a change of terms and the substitution of "dominant vs. recessive" for "unmarked vs. marked" seems to be expedient in view of the ambiguity of the traditional expression "markedness" and its still widespread use in the original sense of "merkmalhaltig/merkmallos" (as different from that of "common, natural" vs. "less common, less natural").

It is precisely the establishment of such universal patterns of compatibility of distinctive features into simultaneous bundles or into "vertical sequences", with determination of their opposite function of "dominance" in the paradigmatic system that appears to be one of the basic tasks of present-day typological phonology. This will help establish universally relevant hierarchical dependence between the correlative units of a phonological system and to identify the core of phonemic oppositions, a kind of deep phonological structure, that constitute the basis of the phonemic inventory of human language, invariant in relation to particular phonemic systems in synchrony and to possible phonemic transformations in diachrony.

In this respect correlations of stops and fricative phonemes in a phonemic system present a special interest. In particular, in the subsystem of stops the following hierarchical correlations of dominance may be established among the phonemic units of various series (cf. Melikishvili, 1970):

In systems with an opposition among stops differing on the feature "voice", the voiced labial /b/ is functionally stronger (dominant) as compared to the velar stop /g/. Stated otherwise, the feature "labiality" in a simultaneous combination with the feature "voice" yields a dominant bundle of features, making up the labial phoneme /b/, as different from the combination of the

features "voice" and "velarity" that yield a functionally weaker, less common and in this sense "recessive" voiced velar stop /g/. Inversely, in the class of voiceless stops it is precisely the velar /k/ that appears as a more natural, functionally stronger and dominant member of the paradigmatic opposition as compared to the labial /p/ serving as a functionally weaker, recessive unit. Thus "velarity" combined with "voicelessness" and "labiality" combined with "voice" form more natural and common bundles of features representing the dominant phonemes /k/ and /b/, whereas the combinations of "voicelessness" with "labiality" and of "voice" with "velarity" create functionally weak, recessive units /p/ and /g/, this being due to the acoustic-articulatory characteristics of the features involved.

Gaps in the paradigmatic system are distributed according to the established functional correlation of dominance of the phonemic units. Systems with gaps in the class of stops opposed according to "voice/voicelessness" assume in general the form as in (1-3):

(1) b -	(2) b p	(3) b -
d t	d t	d t
g k	- k	- k

The degree of recessiveness in the class of voiceless stops increases in accordance with the superposition on the bundle of the additional feature "aspiration" or "glottalization"; incidentally, "glottalization" appears as a feature of a higher degree of "recessiveness" than does "aspiration", so that the hierarchical sequence of increasing dependence in the class of unvoiced stops has the form: voiceless (plain) - aspirated - glottalized. Thus, the glottalized labial /p^h/ appears in relation to the aspirated /p^h/ as a recessive member of the opposition, whereas the aspirated /p^h/ is recessive in relation to the voiceless plain phoneme /p/ (cf. Greenberg, 1970).

Gaps in the paradigmatic systems are represented in accordance with these correlations. The possible systems with gaps in the respective series of voiceless stops are given in (4) and (5):

(4) b p ^h -	(5) b - -
d t ^h t'	d t ^h t'
g k ^h k'	g k ^h k'

There appears to be a further dependence in the paradigmatic system between the subclass of stops and that of the corresponding fricative phonemes which manifest analogous correlations of dominance.

In the labial series the voiced fricative phoneme w/v/β emerges as the dominant member of the correlation, with the recessive voiceless unit /f/, whereas in the velar series the voiceless fricative /x/ functions as the dominant unit as opposed to the recessive voiced fricative /ɣ/, i.e., f → w/v/β, γ → x, and γ → w/v/β, f → x (where the arrow is directed from the recessive member of the opposition to the dominant one). Systems with gaps in the class of non-strident labial and velar fricatives with an opposition of "voice/voicelessness" assume in general the form as in (6-8):

(6) w/v f	(7) w/v -	(8) w/v -
- x	γ x	- x

The subsystem of fricatives appears in the paradigmatic system as a kind of substitute for the corresponding stops. In particular, the absence in the subsystem of stops of its functionally weak, recessive members (i.e., of the velar phoneme in the voiced series and/or the labial phoneme in the voiceless series) presupposes the presence in the paradigmatic system of the corresponding fricative phonemes (i.e., of the velar fricative in the voiced series, and/or the labial fricative in the voiceless series): $\bar{g} \rightarrow \gamma$, $\bar{p} \rightarrow f$. Thus, the fricative phonemes /f/ and /ɣ/ and the dominant members implied by them, viz. w/v and /x/, respectively, emerge as substitutes for the corresponding stops /p/ and /g/, compensating, as it were, for their absence and thus establishing an equilibrium in the paradigmatic system. It may be asserted that the tendency to such an equilibrium in the system is due to the natural phonetic tendency to a symmetric filling of the three main articulatory zones - labial, dental, and velar - with sounds of consonantal articulation: stops or fricatives. If the system has the recessive stops /p/ and /g/, the presence of their substitutes in the form of the corresponding fricatives /f/ and /ɣ/ is optional. Such phonemes appear in the paradigmatic system as redundant consonantal elements, subject to diachronic changes.

Language systems evince a definite hierarchical order among diverse types of structural, in particular phonological, oppositions indicating the existence of a strict stratification of phonological values. It is in conformity with such universally valid correlations that diachronic phonemic transformations occur in a language. This gives a clue helping us to better understand language change in diachrony and to propose linguistically more realistic and plausible schemes of language reconstruction.

The classical Indo-European comparative grammar deals with a system of Proto-Indo-European stops that appears to be linguistically improbable and unrealistic since it runs counter to the typologically established phonological universals concerning the nature of the system of stops, with different phonemic series and a definite distribution of gaps. This necessitates a total revision of the traditionally postulated three-series-system of Proto-Indo-European stops - I: voiced II: voiced aspirates III: voiceless (with an absent or weakly represented voiced labial /b/) and its reinterpretation as I: glottalized II: voiced aspirates III: voiceless aspirates (with an absent, resp. weakly represented, glottalized labial /p'/), cf. Gamkrelidze-Ivanov (1973); Hopper (1973):

<u>Traditional System</u>				<u>Revised System</u>		
I	II	III		I	II	III
(b)	b ^h	p		(p')	b ^[h]	p ^[h]
d	d ^h	t		t'	d ^[h]	t ^[h]
g	g ^h	k	⇒	k'	g ^[h]	k ^[h]
.
.
.

Such a reinterpretation of the traditional system of Proto-Indo-European stops brings it in full conformity with typological evidence, both synchronic and diachronic, and allows to envisage a more realistic and linguistically plausible picture of Proto-Indo-European.

The evidence of the modern linguistic typology and the theory of language universals in effect necessitates a revision of the traditional schemes of the classical comparative linguistics by ad-

vancing new comparative-historical reconstruction.

This is one of the more practical aspects (finding its application in diachronic linguistics) of the modern linguistic typology and the theory of language universals.

References

- Gamkrelidze, Th. V. (1975): "On the correlation of stops and fricatives in a phonological system", *Lingua* 35.
- Gamkrelidze, Th. V. and V.V. Ivanov (1973): "Sprachtypologie und die Rekonstruktion der gemeinindogermanischen Verschlüsse", *Phonetica* 27.
- Greenberg, J. H. (1970): "Some generalizations concerning glottalic consonants, especially implosives", *IJAL* 36.
- Hopper, P. J. (1973): "Glottalized and murmured occlusives in Indo-European", *Glossa* 7.
- Jacob, F. (1977): "The linguistic model in biology", in *Roman Jakobson. Echoes of his Scholarship*, D. Armstrong and C.H. van Schooneveld (eds.), Lisse: The Peter de Ridder Press.
- Melikishvili, J. G. (1970): "Conditions of markedness for the features of voice, voicelessness, labiality, and velarity", *Matsne* 5.

APPARITION ET DISPARITION DES OCCLUSIVES SONORES PRÉGLOTTALISÉES

André-Georges Haudricourt, Centre National de la Recherche Scientifique, 15, quai Anatole France, 75700 Paris, France.

Quelles sont les conditions linguistiques d'apparition des occlusives sonores préglottalisées? A la différence d'une occlusive sourde ou d'une spirante sonore, une occlusive sonore ne peut pas se prolonger indéfiniment. L'air qui passe à travers le larynx en produisant la sonorité est arrêté ensuite par l'occlusion buccale, de sorte qu'au bout d'un instant, la pression de l'air situé entre le larynx et l'occlusion buccale augmente et devient égale à celle de l'air des poumons et de la trachée-artère; de ce fait, l'air ne passe plus : la vibration laryngale s'arrête. Ainsi, la consonne sonore longue tend à s'assourdir. Pour maintenir la distinction pertinente, il faut diminuer la pression de l'air entre le larynx et l'occlusion buccale, c'est-à-dire fermer le larynx au début (préglottalisation), puis le faire descendre pendant la tenue de l'occlusion buccale. Le caractère injectif de la consonne n'en est que la conséquence, lorsque la désocclusion buccale se produit et que l'espace supraglottal a encore une pression inférieure à la pression atmosphérique.

Les langues indo-aryennes de la vallée de l'Indus, sous l'influence probable d'un substrat dravidien, ont transformé tous leurs groupes de consonnes en gémées; la pertinence phonologique d'une distinction entre simples et gémées, entre sourdes et sonores, avait un rendement important, et l'apparition des préglottalisées s'explique. Ces consonnes ne sont conservées actuellement qu'en sindhi, car c'est seulement dans cette langue que les sonores ordinaires sont réapparues assez tôt pour maintenir l'opposition¹.

En Indonésie, on constate l'apparition de la préglottalisation comme réalisation d'occlusives sonores gémées dans certaines langues, tel le samal, parlé dans l'archipel Sulu des Philippines, ou en bougui, parlé à Sulawesi (Célèbes)².

Les langues miao qui ont pénétré en Indochine depuis un siècle ont des occlusives latérales (mais pas de groupes de consonnes); or les langues indigènes d'Indochine n'ont pas ce type de consonnes. Dans le dialecte meo blanc, cette occlusive latérale a été considérée comme un groupe combinant occlusion et sonorité et est devenue une occlusive préglottalisée. Or, les préglottalisées

sont fréquentes dans les langues indigènes, ce qui a dû favoriser ce changement.

En vietnamien, les deux occlusives sourdes p et t sont devenues préglottalisées sonores au cours du Moyen Age, sans qu'aucune raison linguistique puisse être avancée. Il s'agit ici de causes ethnosociologiques; au cours du millénaire d'occupation chinoise, les anciennes préglottalisées austroasiatiques ont disparu (en devenant nasales) en vietnamien proprement dit, mais ont été conservées au voisinage (par exemple en müöng). Lorsque le Vietnam devint indépendant au X^{ème} siècle, les préglottalisées réapparurent. Le même phénomène eut lieu en khmer : les p et t en contact avec la voyelle accentuée se sont préglottalisés, d'où la valeur donnée à ces lettres dans l'écriture thai dès le XII^{ème} siècle.

Dans l'île de Hai-nan, deux langues introduites au Moyen Age – une langue thai, le bê, et un dialecte chinois min, le hainanais ou hoklo – ont subi cette même mutation des p-, t-³. Le même phénomène est signalé dans les dialectes yüe du sud-est du Guangxi⁴.

En résumé, en Indochine, les langues austroasiatiques et thai ont dû, au cours de leur évolution vers la monosyllabisation, engendrer des groupes initiaux de consonnes qui ont abouti linguistiquement à former des occlusives sonores préglottalisées (c'est ce qu'on constate dans une langue de Formose⁵), puis dans cette aire les langues qui venaient d'ailleurs en ignorant ces consonnes, ou celles qui historiquement les avaient perdues, les ont acquises par influence ethnosociologique. C'est le cas des langues karen : sgaw et pwo, langues tibéto-birmanes ayant pénétré dès le haut Moyen Age dans le domaine des langues austroasiatiques, et thai. Il y a passage de p-, t- aux sonores préglottalisées.

Actuellement, le thai de Bangkok a transformé ses préglottalisées en sonores ordinaires, peut-être au contact des langues européennes, mais les anciennes sonores historiques s'étant assourdies, la glottalisation n'avait plus de pertinence phonologique.

Par contre, lorsque les langues thai arrivent sur le domaine des langues tibéto-birmanes qui ignorent les préglottalisées, ces consonnes tendent à perdre leur occlusion en devenant des nasales préglottalisées (stade attesté par le ton, pour le khamti, le tay-nüa, le zhuang de Po-ai) qui sont maintenant des nasales ordinaires. Dans d'autres dialectes, elles deviennent des spirantes sonores (shan, tay-noir).

Le passage aux sonores ordinaires a dû se produire en khasi, langue austroasiatique isolée en domaine tibéto-birman.

Le penjabi, langue indo-aryenne voisine du sindhi, a dû passer par le même stade que celui-ci, car les occlusives sonores de cette langue sont liées au ton haut; elles ont été préglottalisées. Et lorsque les anciennes sonores se sont assourdiées (comme en khmer), la préglottalisation, n'étant plus pertinente, a pu disparaître.

Références bibliographiques

- (1) Haudricourt, A.-G. (1977): "La préglottalisation des occlusives sonores", Bulletin de la Société de Linguistique de Paris, 52, 1, pp. 313-317, Paris: Klincksieck.
- (2) Reid, L.A. (ed.) (1971): Philippine Minor Languages: Word Lists and Phonologies, p. 34, Oceanic Linguistics special publications n° 8 (256 p.), Honolulu: The University Press of Hawaii.
 Reid, L.A. (1973): "Diachronic Typology of Philippine Vowel Systems" in Current Trends in Linguistics, 11, Diachronic, Areal and Typology, T.A. Sebeok (ed.), xii-604 p., The Hague: Mouton.
- Sirk, J.K. (1975): Bugijskij jazyk, p. 29, Moscou: édition "Nauka", 112 p.
- (3) Haudricourt, A.-G. (1959): "How History and Geography can explain certain phonetic developments", Yǔyán-yánjiu, 4, 81-86, Pékin.
 Hagège C. et Haudricourt A.-G. (1978): La phonologie panchronique, Paris: Presses Universitaires de France.
- (4) Tsuji, N. (1977): "Murmured Initials in Yue Chinese and Proto-Yue Voiced Obstruents: the Case of Cenxi Dialect, Guangxi Province", Gengo Kenkyu, 72, 29-46, Tokyo.
- (5) Tsuchida, S. (1972): "The Origins of the Tsou Phonemes /b/ and /d/", Gengo Kenkyu, 62, 24-35, Tokyo.

TYPOLOGICAL UNIVERSALS, ASPIRATION, AND POST-NASAL STOPS

Robert K. Herbert, Department of Linguistics and Oriental and African Languages, Michigan State University, East Lansing, MI 48824, U.S.A.

Introduction

Probably the least marked type of consonant cluster found among the world's languages is Nasal + Oral Consonant (NC). The unmarked status of this sequence is demonstrated by a number of factors, including their occurrence in many languages otherwise characterized by CVCV structure. Perceptually, such a sequence is easily exploited since nasal consonants, although easily confused within the class, are quite distinct from oral consonants. The confusion within the class accounts, in part, for the fact that NC sequences are very frequently homorganic. Articulatorily, the sequence of gestures required to produce a NC cluster is relatively simple, involving only a raising of the velum for the sequence nasal plus voiced stop (ND).¹ For other types of NC clusters other gestures are necessary such as a cessation of vocal fold activity and a reduction in the degree of stricture. Further, the optimal opposition within NC sequences is demonstrated by its frequent exploitation in unit sound types, the so-called "half-nasal consonants", pre- and postnasalized consonants.²

The degree of articulatory and perceptual complexity is mirrored in the relative markedness of NC types. Thus, the least marked type of cluster is ND. Other types occur, even among the half-nasals, but these are less common³ and many derivational processes, both synchronic and diachronic, conspire to produce NC inventories of the least marked type.

-
- (1) The following symbol abbreviations will be used within the text: Nasal + Voiced Stop (ND), Nasal + Voiceless Stop (NT), Nasal + Voiced Fricative (NZ), etc. Other symbols employed have their standard phonetic values.
 - (2) The half-nasal consonants are distinguished from NC clusters by a number of factors, the most essential being that of duration. The two components of a half-nasal will exhibit the combined surface duration equivalent to a single consonant.
 - (3) This frequency is demonstrated in both cross-language frequency of occurrence and text frequency. In a 1000 phone count of a Rundi text, the following statistics were obtained: NC (30): ND (21), NT (4), NZ (4), NS (1).

A typological survey of the processes affecting either component of a NC sequence provides two inventories of process, one affecting the nasal and one the oral consonant. Among the former, only homorganicity assimilation is common whereas positional assimilation of the oral consonant is rare. Perhaps the most common process affecting the oral consonant is post-nasal voicing of voiceless consonants. In such a sequence, there are two primary motions which distinguish the two components: (1) raising of the velum, (2) cessation of vocal fold vibration. If the two are not coordinated the following sequences obtain: (a) NÇÇ, or (b) NŇÇ. In many languages the former tendency has been phonologized so that all post-nasal consonants are voiced.

Another common process is post-nasal hardening which, in conjunction with voicing, accounts for some of the many inventories containing only ND. Hardening actually involves two subtypes, but since many languages exhibit these in conjunction, it is perhaps best to view this situation as a continuum:

continuant → affricate → stop

In many cases, the hardening effect of nasals is evident even after the nasal is lost historically.

Other processes not of relevance to the present paper include post-nasal de-implosion (Shona /N+ð/→[mb]), ejectives (Zulu /N+ph/→[mp?]), etc. The situation with regard to aspiration of voiceless stops is problematic. On the one hand, some languages exhibit clear patterns demonstrating the loss of aspiration in this environment. However, other languages show aspiration developing in this context. Thus, there are conflicting tendencies which exist with regard to aspiration. This is not a felicitous situation since it is otherwise possible to determine a general direction of evolution. While changes of the sort NÇ → NŇ, NT → NTS occasionally occur, they are rare and other factors are found which explain these anomalous developments.

Loss of Aspiration

In Zulu, aspiration is lost in contact with nasal consonants. Doke (1926) reports the development of ejectives from aspirates in this context, but not all speakers exhibit this tendency. Aspirated clicks are replaced by simple nasal clicks when they are brought under nasal influence in Zulu whereas they merely lose

their aspiration in Xhosa. Tarascan (Foster 1969) has two series of underlying non-nasal obstruents /p t c č k/ and /p^h t^h c^h č^h k^h/. In contact with nasal consonants, the plain consonants are voiced, and the aspirates become plain voiceless consonants, e.g. /N+p/ → [mb], /N+p^h/ → [mp]. Devine (1974:19) notes that it may be best to regard this as a sliding scale of complexity and that the normal state for voiceless consonants in contact with preceding sonorants is unaspirated.

Development of Aspiration

In his useful survey of the noun class system of Bantu, Kadima (1969:63-5) notes that the most common developments of NT sequences are:

$$\begin{aligned} /N + p t k/ &\rightarrow [p t k] \\ &[p^h t^h k^h] \\ &[mp^h nt^h \eta k^h] \end{aligned}$$

Other developments also occur, e.g. [mb nd ŋg], [m̄ ŋ ŋ̄]. The present concern is with the development of aspiration. In Venda (Ziervogel and Dau 1961), Bantu nasal compounds develop as follows:⁴

*mb > mb	*mp > p ^h
*nd > nd	*nt > t ^h
*ŋg > ŋg	*ŋk > k ^h

When the simple stops are not under nasal influence, they undergo spirantization:

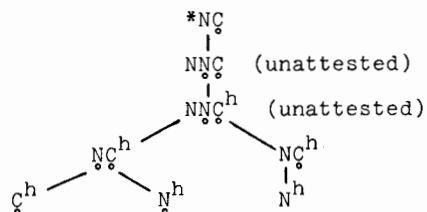
$$\begin{aligned} *p t k &> \phi r h \\ *b d g &> \beta l \emptyset (j) \end{aligned}$$

However, not all languages which develop aspiration exhibit a weakening of stops otherwise; it is therefore not possible to attribute aspiration to any general weakening process.

Hinnebusch (1975) attempted to reconstruct the phonetic processes in Swahili by which *mp nt ŋk became p^h t^h k^h. He proposed a two-stage process, the first of which is nasal devoicing, followed by deletion. It is proposed that native speakers reinterpreted the period of initial noisiness as post-aspiration rather

(4) However, the nasal is retained in both series with monosyllabic stems although it comprises a separate syllable: ŋk^h 'large pot', nt^h 'louse'. Further, the nasal is retained if it represents the first person singular object marker.

than preaspiration. The following is attributed to John Ohala:



Givon (1974) explains the development of aspiration by reference to three facts

- i) assimilatory devoicing of nasals before voiceless stops
- ii) voiceless nasals tend "to be perceived as breath"
- iii) voiceless stops tend to be universally aspirated

A perceptual confusion arises and there is a metathesis in which nasal breath is interpreted as post-aspiration. Ignoring for the moment the assertion that aspiration is the natural state for voiceless consonants universally, a universal of doubtful validity, the metathesis analysis seems plausible.

The two unattested stages above represent periods of variation before any phonetic tendency is phonologized. Wide variation in the realizations of NT sequences are found in many languages, e.g. in Malagasy /mp/ may be [mp, mp̚, hp, p, pʰ, ph].

Further Development of the Aspirates

Aspirated stops derived in this manner are liable to other developments after the nasal has been lost. Frequently, they develop into fricatives or affricates. There is much comparative evidence to support this, e.g. Tswana *mhaxo*, Pedi *mp^haY_o*, Sutho *mofao* 'provisions'. (Cf. also the development of postnasalized stops into aspirates and fricatives in many New Caledonian languages (Haudricourt 1964, 1971).) Languages frequently pass through an affricate stage before the fricative inventory is established. Hyman (1974) argues that even when there is no evidence for such a stage we may assume a "telescoping" of process. The important point here is to note that these developments occur only after the nasal has been lost; this explains why correspondences such as *mp nt ŋk → f θ x do not violate the universal of hardening discussed above. Similarly, Sango *ŋk > ŋx must have passed through an intermediate stage *ŋkx (<*ŋkh) which derives from aspiration being interpreted as a velar fricative due to the

acoustic similarities between the two. This seems plausible when we view the rest of the NT series: *mp̚ > mh, *nt > nh. In fact, the aspiration of velars in many languages is often phonetically [x], e.g. Scots Gaelic (Ternes 1973).

Another seemingly anomalous situation is presented by languages in which stops are voiced except after a nasal. This is certainly a preferred environment for voicing, yet there are correspondences such as Bulu:

*p > v	*mp > f
*t > l	*nt > t
*k > Ø	*ŋk > k

It is necessary to explain the non-voicing of /f t k/ as resulting from previous aspiration, which prevented voicing in this position.

A complete inventory of processes affecting the derivation of NC sequences is beyond the scope of this short paper. Although there are relatively few processes which operate on ND sequences, really only simplification in favor of the oral or nasal consonant, a number of processes conspire to produce NC inventories which include only ND sequences. These include both direct and indirect processes, i.e. those which change feature specifications and those which eliminate one component of the sequence. Apart from the universal primacy of ND sequences, there may be language-specific variation in terms of the relative weightings of other types, e.g. NT, NZ.

Typology and Reconstruction

Part of the value of surveys of evolutionary processes is that they serve as useful tools in diachronic linguistics. This idea is far from novel. Jakobson (1958) noted that such studies form the touchstone of validity for all reconstructed systems. The interaction of processes of change as well as the directionality of change itself often provide insight into problems of reconstruction. Studies of this sort point not only backwards to possible sources of origin, but also forwards to future directions of possible change.

Bennett (1967), in discussing the voicing of post-nasal stops in several Eastern Bantu languages, reconstructs the phonetics of change as:

*mp > *mɸ > *mβ > mb

*nt > *nθ > *nð > nd

Nasality is lost in certain cases and *mp > ɸ or f. However, there are several serious problems with the proposed reconstruction. Specifically, the change *mp > mɸ is unlikely to the extent that consonants tend to harden in this environment. In fact, the sequences [mɸ mβ nθ nð] are all uncommon. Although [mβ] occurs, it always represents /mβ/, never /mb/, and the more common realizations of such a sequence are [mb, β, b]. Ladefoged (1968:47) reports the existence of [nθ] in Sherbro, a surprising fact since Sherbro also exhibits /f v s/, none of which appear after a nasal. Kamba exhibits [nð]. On the whole, however, this is a restricted class of sounds.

Further, the fact that intervocalic voiceless stops lenite cannot be cited as evidence that post-nasal stops behave similarly. There are numerous examples where the two develop differently. For example, Londo *mp nt ŋk > p t k whereas p t k > β t x. In Mbɔle, *mp nt ŋk > f t k and *p t k > ɸ t ø. A crucial fact in cases exhibiting the development of a fricative from a voiceless stop is that nasality is lost. In such a case, intermediate stages are attested elsewhere, e.g. Lwena *mp nt ŋk > p^h t^h k^h and p t k > h t k. Also, the existence of nasal and fricative series generally implies the existence of nasal and stop series, which condition is not met by Bennett's system. Thus, the proposed reconstructed chronology cannot be accepted, especially in view of the frequency and naturalness of the process whereby consonants are voiced after a nasal consonant.⁵ The point here is that although it is necessary to make inferences about the phonetics of prehistory, these inferences must be solidly grounded in a theory of universal processes and phonetics. There are definite limitations to be placed upon the importance attached to such studies for other purposes,

(5) Cases such as Makua *mb nd ŋg > p t k must involve two distinct stages: (1) nasal loss, (2) later devoicing. There is no neutralization of NC series since *mp nt ŋk > p^h t^h k^h. One step neutralizations of NC series always favor the voiced series, e.g. Yao *mp, mb > mb; *nt, nd > nd; *ŋk, ŋg > ŋg.

e.g., genetic classification, linguistic subgroupings, etc.

Conclusion

This brief paper has attempted to demonstrate how various claims made by Jakobson, Greenberg, and others may be applied to the study of NC sequences. This included an examination of the relationship between synchronic universals and diachronic processes and between typology and universals. Greenberg (1970a:61) points out that the former follows logically from the fact that no change can produce a synchronically unlawful state and that all states are the outcome of diachronic processes. The distinction between state and process is an important one. The general direction of NC evolution toward the least marked ND sequence again supports the generalization that diachronic process explains frequency in phonology. The predictive power of typological studies demonstrates this complex interaction between the shape and patterning of phonological systems.

References

- Bennett, P. (1967): "Dahl's Law and Thagicũ" African Language Studies 8, 127-59.
- Devine, A. (1974): "Aspiration, universals, and the study of dead languages", Working Papers in Language Universals 15, 1-24.
- Doke, C. (1926): The Phonetics of the Zulu Language, Bantu Studies, Special Number.
- Foster, M. (1969): The Tarascan Language, Berkeley: UCPL 56.
- Givon, T. (1974): "Rule un-ordering: generalization and degeneralization in phonology", Papers from the Parasession on Natural Phonology, 103-15, Chicago: Chicago Linguistic Society.
- Greenberg, J. (1969): "Some methods of dynamic comparison", in Substance and Structure of Language, J. Puhvel (ed.), 147-203, Berkeley: University of California Press.
- Greenberg, J. (1970a): "Language Universals", in Current Trends in Linguistics, T. Sebeok (ed.), 3,61-112. The Hague: Mouton.
- Greenberg, J. (1970b): "The role of typology in the development of a scientific linguistics", in Theoretical Problems of Typology and the Northern Eurasian Languages, L. Dezső and P. Hajdú (eds.), 11-24, Bucharest: Akadémiai Kiado.
- Guthrie, M. (1967-70): Comparative Bantu, 4 vols., Farnborough: Gregg International Publishers.

26 SYMPOSIUM No. 1

- Haudricourt, A. (1964): "Les consonnes postnasalisées en Nouvelle Calédonie", Proc. Ling. 9, 460-61, The Hague: Mouton.
- Haudricourt, A. (1971): "Consonnes nasales et demi-nasales dans l'évolution des systèmes phonologiques", Proc. Ling. 10, 4, 105-8, Bucharest: l'Académie de la République.
- Herbert, R. (1977): Language Universals, Markedness Theory, and Natural Phonetic Processes: The Interactions of Nasal and Oral Consonants, Unpublished Ph.D. dissertation, Ohio State University.
- Hinnebusch, T. (1975): "A reconstructed chronology of loss: Swahili class 9/10", in Proc. of the Sixth Conference on African Linguistics, R. Herbert (ed.), 32-41, Columbus: OSUWPL 20.
- Hyman, L. (1974): "Contributions of African linguistics to phonological theory", Fifth Conference on African Linguistics, Stanford University. Ditto.
- Jakobson, R. (1958): "What can typological studies contribute to historical comparative linguistics?", Proc. Ling. 8, 17-35, Oslo: Oslo University Press.
- Kadima, M. (1969): Le système des classes en bantou, Leuven: Vander.
- Ternes, E. (1973): The Phonemic Analysis of Scottish Gaelic. Hamburg: Helmut Buske Verlag.
- Ziervogel, D. and R. Dau (1961): A Handbook of the Venda Language, Pretoria: University of South Africa.

UNIVERSALS OF VOWEL SYSTEMS: THE CASE OF CENTRALIZED VOWELS

Jean-Marie Hombert, Linguistics,
University of California, Santa Barbara, USA 93106

This paper attempts to explain why centralized vowels (i.e. vowels which are not located on the periphery of the vowel space) are relatively less common than peripheral vowels.

1. Surveys of phonemic systems, phonetic universals and "exotic" languages.

If one is interested in discovering phonetic universals some of the most fruitful places to search for potential universals are large scale surveys of phonetic and phonemic inventories. Despite the criticism leveled against these surveys it is our belief that such surveys are useful in that asymmetries or systematic gaps in these inventories may reveal in their explanation universal phonetic processes. Once such a potential universal or universal tendency has been uncovered each language exhibiting this process should be reexamined through careful study of available sources, consideration of possible reinterpretations of the data, and when possible, accurate phonetic data should be obtained.

Until very recently the bulk of available phonetic data, especially perceptual data, has come from a handful of languages. Due to the availability of phonetic equipment and presence of research groups located in the countries where these languages are spoken available phonetic data has been largely limited to Danish, Dutch, English, French, German, Japanese and Swedish. It is clear that if we are to understand universal phonetic processes, our data base must be extended to include more "exotic" languages.

Most perceptual data has been gathered from experiments conducted under laboratory conditions using linguistically sophisticated subjects. Obviously if we are to gather similar data from languages spoken in areas remote from laboratory facilities, it is necessary to design techniques of data gathering suitable for use in the field with linguistically naive subjects. In Section 3 one such design will be discussed.

2. The case of centralized vowels.

It is clear from surveys of vowel systems that centralized vowels are less commonly found than peripheral ones. In the case of languages which do have centralized vowels it is not rare that different sources will vary in the treatment of such vowels by

attributing to a given vowel different phonetic qualities. These variations suggest that either these vowels are more prone to historical change or are more difficult to identify correctly by the investigator. It appears, then, from these surveys that non-peripheral vowels, that is, vowels which in acoustic terms have a second formant of approximately 1200-1700 Hz, are rare and that they are more subject to change than peripheral vowels.

In Section 3 we will use data from a perceptual experiment carried out on the Grassfield Bantu languages of Cameroon. Because of space constraints in this paper, we will use only data from one speaker of the Fe?fe? language¹ to suggest possible explanations for the rarity as well as instability of non-peripheral vowels.

3. Experimental paradigm

Fe?fe? contains eight long vowels in open syllables. These vowels are [i, e, a, v, o, u, w, ə]. A word list consisting of eight meaningful Fe?fe? words contrasting these eight vowels was elicited from native Fe?fe? speakers. The Fe?fe? speakers were asked to read these eight words which were listed five times each, in random order. After the repetition of each word, the final sound of the word, that is the vowel, was repeated once. Both the vowels of the meaningful words and the vowels in isolation were subsequently analyzed.

Subjects were then asked to listen to 53 synthetic vowel stimuli, each presented five times in random order. After the presentation of each stimulus the subjects were instructed to point out which Fe?fe? word in the eight-word list that they had previously read, contained the same "final sound", i.e. vowel, as the stimulus. Subjects had the option to claim that some of the stimuli did not sound like any of the eight Fe?fe? words. The 53 synthetic stimuli were selected to maximally cover the vowel space; F1 was varied between 250 Hz-750 Hz, F2 between 650 Hz-2350 Hz and F3 between 2300 Hz-3100 Hz. This task was designed so that native speakers would divide the vowel space according to their own vowel systems.²

4. Results

The results of the acoustic analysis and of the perceptual

-
- (1) For more data and a more complete description of the experimental paradigm, see Hombert (in preparation).
 - (2) It should be noticed that this method does not allow study of diphthongs since all stimuli used have steady state formant frequencies.

experiment for one Fe?fe? speaker are presented in Figure 1 and Figure 2 respectively. Since F3 values are not relevant for the point that we want to make here the data are presented in an F1 x F2 space. Each vowel indicated in Figure 1 is the average of five measurements. The spectra were computed 100 msec. after vowel onset using LPC analysis. The phonetic symbols appearing in Figure 2 indicate that at least four times out of five this stimulus was identified by the Fe?fe? speaker as the same vowel.

We will consider the two vowels [a] and [ə]. Two unexpected results emerge from the data:

1. When comparing acoustic and perceptual data it is not surprising to find that the stimulus with F1 at 750 Hz and F2 at 1250 Hz is identified as the vowel [a] since a vowel with such a formant structure could have been produced by a Fe?fe? speaker with a larger vocal tract size than the speaker considered here. What is surprising, though, is that the stimulus with the formant structure F1 at 750 Hz and F2 at 850 Hz was also identified as [a]. These results are even more surprising when one considers that the intermediate stimulus (750 Hz - 1050 Hz) was identified as [v]. It is likely that in the case of the stimulus with F1 at 750 Hz and F2 at 850 Hz the two formant peaks were perceived as one formant peak, that is as F1. One thing remains to be explained: in the acoustic data, the Fe?fe? vowel [a] has a peak around 1600 Hz but the stimuli with F1 at 750 Hz and F2 at 850 Hz does not have a peak in this frequency region. Let us just say for the moment that the saliency of the peak at 1600 Hz seems to be perceptually secondary.
2. Two stimuli (F1 at 350 Hz, F2 at 1500 Hz and F1 at 450 Hz, F2 at 1500 Hz) are identified as [ə], which is what we would expect considering the location of [ə] in Figure 1. However the identification of the stimulus with F1 at 450 Hz and F2 at 650 Hz with [ə] comes as a surprise. Notice that F1 and F2 are also close to each other for this last stimulus, which could have lead to the perception of them as one peak corresponding to the first formant. But notice also that this stimulus does not have a peak around 1500 Hz. As in the case of the vowel [a] it appears that the perceptual saliency of the peak around 1500 Hz did not play a major role in the identification of the [ə].

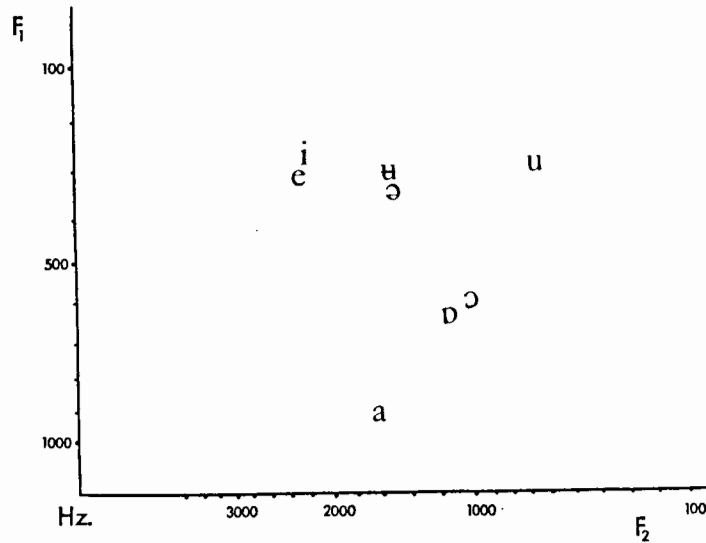


Figure 1. Acoustic data: the Fe?fe? vowel system, (one speaker, average of five measurements).

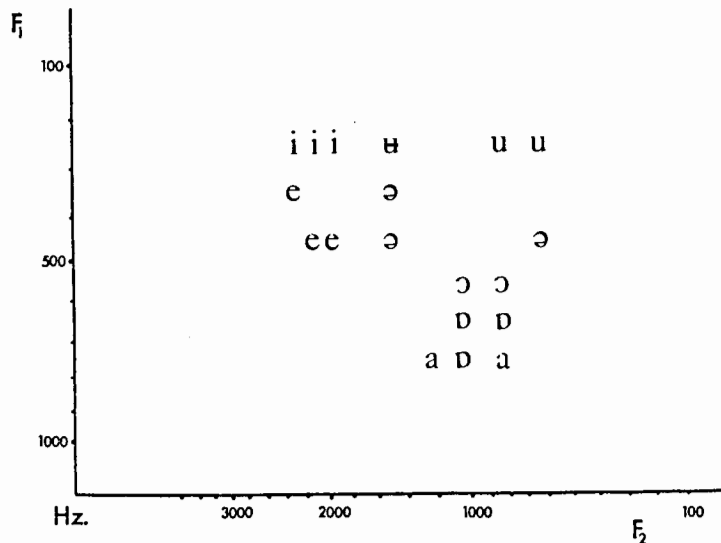


Figure 2. Perceptual data: only stimuli for which the Fe?fe? subject gave at least four out of five identical responses are presented on this graph.

5. Discussion³

Two possible explanations to account for the lack of saliency of formant peaks around 1500 Hz are being explored now.

1. Spectrum-based representation of vowels.

Our results would be compatible with a mechanism of vowel perception which looks for certain amounts of energy within frequency regions rather than formant peaks. In the cases which we discussed in the previous section, the unexpected vowel identification happened with stimuli which had their first and second formants very close to each other. In such cases the closeness of the first two peaks leads to an increase in amplitude of the spectrum. This increased amplitude may have created sufficient energy in the 1500 Hz region to lead to these "perceptual mistakes".

2. Place vs. periodicity mechanisms.

Pitch is processed by different mechanisms depending upon its frequency region. The boundary between these two mechanisms (place vs. periodicity) is not well defined. It is possible that for some subjects a defective overlap between these two mechanisms in the 1500 Hz region could create the perceptual mistakes presented in Section 4.

6. Implications

The explanation generally provided to account for the relative scarcity of non-peripheral vowels is based on the principle of maximum perceptual distance presented by Liljencrants and Lindblom (1972). Our results suggest a different explanation - non-peripheral vowels are avoided because one of their components (F2) is located in a relatively less salient perceptual zone. If this is the case we can now understand why processes leading to vowel centralization (vowel nasalization, rounding of front vowels, unrounding of back vowels) are relatively uncommon.

Finally we should point out that "perceptual mistakes" such as the ones reported in Section 4 were found in approximately one out of five subjects, with the "mistake" being consistently made by the one subject. These results would be consistent with a theory of sound change which claims that sound changes are initiated by a minority of speakers.

(3) The reason why previous experiments on vowel perception did not uncover this problem may be due to the nature of the experimental paradigm as well as the range of stimuli used in this experiment.

- (vii) $\left\{ \begin{array}{l} a \rightarrow \text{ɔ/w} \text{ __} \\ \emptyset \rightarrow \text{æ/ __ r} \end{array} \right.$ (was, swan, quarrel; Middle English).
 ([\emptyset :va] vs [æ :ra]; Swedish).
- (viii) $\left\{ \begin{array}{l} x \rightarrow \left\{ \begin{array}{l} \text{ç / +front V __} \\ x / \text{elsewhere} \end{array} \right. \\ /h/ \text{ realizations of Japanese cf. (v) above.} \end{array} \right.$ ([l₁çt] vs [axt]; German).
- (ix) $\left\{ \begin{array}{l} n \rightarrow m / b \text{ __} \\ /n/ \text{ realizations of Swedish cf. (iii) above.} \end{array} \right.$ ([ha:bm] (haben); German).
- (x) $\left\{ \begin{array}{l} r \rightarrow \text{ʀ} / \left[\begin{array}{l} \text{-voic} \\ \text{-son} \end{array} \right] \text{ __} \\ \text{ʒ} \rightarrow \text{ʒ} / \text{ __ } [-\text{voic}] \\ k \rightarrow \text{k} / \text{ __ } [+ \text{voic}] \end{array} \right.$ ((try, cry, pry; English).
 ([neʒʒfɔ̃dy] French
 [saʔkɔ̃ʀ] French)

The above examples of pro- and regressive assimilations suggest that assimilation be hypothetically described as a reduction of articulatory distance in articulatory space. Do they imply a syntagmatic pronounceability condition, favoring a reduction of the physiological equivalent of a power constraint, mechanical work (force x distance)/time (a LESS EFFORT principle)? Can at least some phonological facts be interpreted as cases of contrast-preserving articulatory simplifications? What is their behavioral origin?

3. Speech - a Physiological Pianissimo.

3.1 The question also arises whether spoken language underexploits the degrees of freedom that in principle the anatomy and physiology of speech production make available. Seen against the full range of capabilities, speech gestures, like many other skilled movements, appear to be physiologically "streamlined" both as regards muscle recruitment and force levels (cf. jaw closure as a speech gesture and in mastication, speech breathing vs respiration in general, articulatory gestures vs swallowing etc.). Extreme displacements of articulatory organs do not occur (PIKE 1943, 150) although such configurations are available and yield acoustically equivalent results (evidence from non-speech: body-arm, eye-head coordination; and from speech: lip/tongue-mandible and tongue blade-tongue body coordination (LINDBLOM et al 1974)). Do we in these circumstances see the operation of an economy of effort principle? A principle that we should invoke to explain how and why speech and non-speech sounds differ

and to account for certain phonological regularities as well as the instances of hypo-articulation (reductions, ellipses, co-articulations etc.) in spontaneous speech. "Today's allophonic variation leads to tomorrow's sound change..." OHALA (1979).

3.2 Pronounceability and Syllable Structure.

FIG. 1 shows average measurements of jaw positions for Swedish apical consonants in the environment [a'Ca:]. The production of these consonants permits a variable influence of the open jaw positions of the vowels. Thus the dimension of jaw opening reveals one aspect of their "willingness" to coarticulate. It is of considerable interest to see that this measure correlates well with their universally favored position in initial and final phonotactic structures (ELERT 1970). If the present observations are generalized, they imply that the phonetic structure of clusters can be explained at least in part with reference to ease of co-articulation (ELERT 1970, BRODDA 1972).

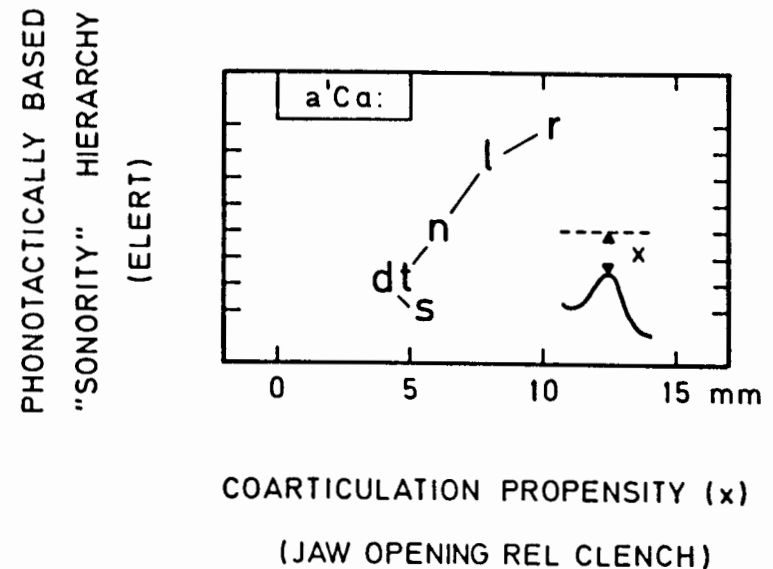


FIG. 1

4. The Distinctiveness "Conspiracy".

4.1 Language structure exhibits redundancy at all levels.4.2 Speech generation is an output-oriented process: The reference input to the speech control system is specified in terms of a desired output. The dimensions of the target specifications are sensory, primarily auditory. Evidence supporting the primacy of auditory targeting comes from work on compensatory articulation, speech development and the psychological reality of phonological structure (LINDBLOM et al to appear, LINELL 1974).4.3 Speech understanding is an active (top-down or conceptually driven) process. (Cf. the demonstrations of context-sensitive processing, resistance to signal degradation, phonemic restoration, verbal transformation etc.)4.4 The speech system may possess specialized mechanisms that contribute towards enhancing the distinctiveness of stimulus cues. Examples of such hypothetical mechanisms are "feature detectors" in speech perception. Specialization of speech production has been suggested in the case of the phylogenetic development of the human supralaryngeal vocal tract whose shape LIEBERMAN (1973) interprets as a primarily speech-related adaptation increasing the acoustic space available for speech sounds.4.5 Phonetic targets are selected so as to retain acoustic stability in the face of articulatory imprecision (STEVENS 1972).

The properties listed in 4.1 through 4.3, do they have a common origin in a basic principle of language design viz., the DISTINCTIVENESS CONDITION: different meanings sound different? The preservation of meaning across encoding and decoding seems to be favored by redundancy, output-oriented and active processing (rather than by lack of redundancy, exclusively input-oriented encoding and purely passive decoding strategies). Thus the question arises whether these at first seemingly unrelated attributes form an evolutionary "conspiracy". Do they constitute three different ways of coping with a physical signal which is inevitably going to be noisy, variable and ambiguous? 4.4 and 4.5 could offer related advantages. What is the behavioral origin of the distinctiveness condition?

5. Speech Development.

5.1 Imperfect learning: Can perceptual similarity and articulatory reinterpretation serve as a source of phonological innova-

tion (cf. JONASSON (1971))? Many sound substitutions in children's speech appear compatible with this interpretation: $\theta \rightarrow f$, $\lambda \rightarrow w$ cf. 2.1. The child is a cognitive and phonetic bottle-neck through which language must pass. Does the process of acquisition leave its imprints on language structure?

5.2 Selection of the fittest: A speech community may use in free variation several realizations of a given form. The set of fricatives may contain /f, s, ʃ, ç/ and /h/ with the /ʃ/ produced as [ʃ] and [ʃ̥] (cf. Swedish). The distinctiveness principle favors [ʃ] which contrasts better with [ç] than [ʃ̥]. The lower confusion risk of the pair [ʃ] / [ç] promotes its reception and learning by the child. There is in this case thus a behavioral rather than teleological motivation for the distinctiveness condition. If sound patterns show evidence of perceptual differentiation, is communicative "selection of the fittest" among several competing forms one of the evolutionary mechanisms? Selection occasionally occurs from a rich variety of hypo- as well as hyper-articulated forms (STAMPE 1972). Is hyperarticulation another behavioral source of distinctiveness?

6. Non-Phonetic Origins of Sound Patterns: Social Biasing.

Selection of speech forms is influenced not only by production and perception factors. Phonological contrasts vary as a function of social variables (prestige, age, class, sex, style etc.). Does the interaction of the sometimes conflicting requirements of social and phonetic factors account for the fact that there is no evidence (GREENBERG 1959) that language change leads to more efficient linguistic systems? Is local rather than global phonetic evaluation of systems (KIPARSKY 1975) another reason why languages do not seem to be converging toward a single optimum equilibrium?

The emergence of a phonological system can be simulated on the basis of current models of production and perception. FIG. 2 shows some computational results obtained by an application of

$$\sum_{i=2}^n \sum_{j=1}^{i-1} T_{ij}(t) \cdot L_{ij}(t) \cdot S_{ij}(t) < \text{CONSTANT} \quad (1)$$

where n is the size of a universal inventory of segments, T_{ij} represents a (time-varying) talker-dependent measure of evaluation for a given contrast (pronounceability condition), L_{ij}

refers to a listener-dependent evaluation (distinctiveness condition), and S_{ij} reflects the balance between social and phonetic factors. FIG. 2 illustrates the

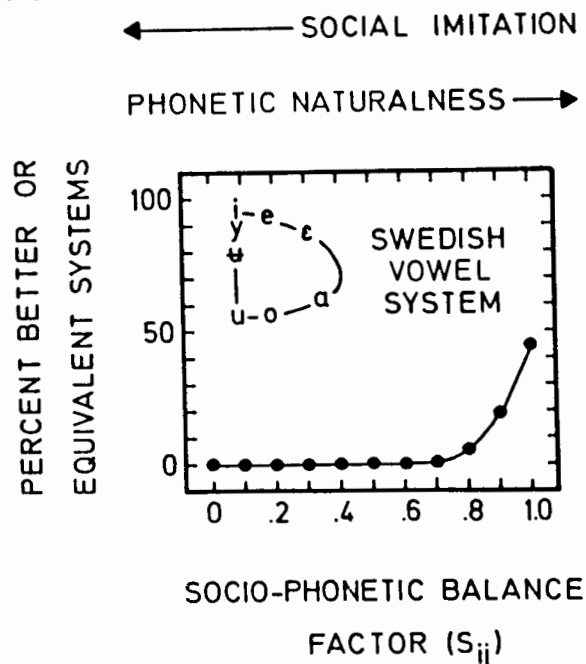


FIG. 2

interaction between the criteria of distinctiveness and social imitation in deriving the Swedish vowel system from a larger set of universal vowel types (represented in terms of canonical auditory patterns). The socio-phonetic balance varies from zero ("social imitation" dominates) to unity (natural phonetic factors, T and L, dominate). It is applied to the contrasts of Swedish with the values shown. For non-Swedish contrasts $S=1$. Apparently there are many systems (out of a total of 92378) that meet our present criterion of distinctiveness equally well or better. If we had reason to believe that the role of natural phonetic factors in the genesis of the Swedish vowels was correctly and exhaustively reflected in our calculations we would conclude that social factors are quite important in their development. We don't. A great deal of work on phonetic naturalness remains to be done before any safe conclusions can be drawn.

However, we believe that the approach will be useful in studying phonological contrasts particularly in child language and cross-linguistically.

7. A "Darwinian" Theory of Phonological Universals.

Suppose that we answer all the questions of the preceding discussion in the affirmative. We accept as our null hypotheses the assumptions that learnability, pronounceability and perceptibility conditions can account for differences between speech and non-speech sounds, that discreteness reflects the operation of memory, learning and decoding mechanisms, that sound changes are influenced by social variables and shaped by demands for perceptual efficiency and convenience of production, and that the origin of such demands is prosaically behavioral rather than mysteriously teleological. Such acceptance boils down to the idea that phonological structure arises both phylogenetically and ontogenetically by "natural selection" of sound patterns from sources of phonetic variation. Language structure emerges in response to the biological and social conditions of language use. Natural selection is based on the communicative (perceptual as well as social) value of contrasts and "phonetic variation" is defined with respect to possible segment, possible sequence and their possible variation. According to this "Darwinian" theory, phonological universals will be explained with reference to how speech is acquired, produced and understood, or rather in terms of our models of these processes.

This conclusion may seem uncontroversial. However, a truly quantitative and explanatory theory along these lines is not likely to appear until we learn to recognize its full intellectual, educational and administrative implications for how linguistics should be done. Language is the way it is partly because of our brains, ears, mouths as well as our minds. In this sense linguistics is a natural science. Phonetics can contribute by formulating its behavioral explanans principles.

8. References.

- BRODDA, B. (1973): "Naturlig Fonotax", unpubl. manuscript, Stockholm University.
- CHAFE, W.L. (1970): Meaning and Structure of Language, Chicago and London: The University of Chicago Press.
- ELERT, C.C. (1970): Ljud och Ord i Svenskan, Stockholm: Almqvist & Wiksell.

- GREENBERG, J.H. (1959): "Language and Evolution", in MEGGERS, B.J. (ed.): Evolution and Anthropology: A Centennial Appraisal, pp. 61-75.
- JONASSON, J. (1972): "Perceptual Factors in Phonology", in RIGAULT, A. & CHARBONNEAU, R. (eds.): Proceedings in the Seventh International Congress of Phonetic Sciences, pp. 1127-1131, The Hague: Mouton.
- KIPARSKY, P. (1972): "Explanation in Phonology", in PETERS, S. (ed.): Goals of Linguistic Theory, pp. 189-227.
- KIPARSKY, P. (1975): "Comments on the Role of Phonology in Language", in KAVANAGH, J.F. and CUTTING, J.E. (eds.): The Role of Speech in Language, pp. 271-280.
- LIEBERMAN, P. (1973): "On the Evolution of Language: A Unified View", Cognition 2 (1), pp. 59-94.
- LINDBLOM, B., PAULI, S. and SUNDBERG, J. (1974): "Modeling Coarticulation in Apical Stops", in FANT, G.: Speech Communication, vol. 1, pp. 87-94, Almqvist & Wiksell Int.
- LINDBLOM, B., LUBKER, J. and GAY, T. (in press): "Formant Frequencies of Some Fixed-Mandible Vowels and a Model of Speech Motor Programming by Predictive Simulation", J. Phonetics.
- LINELL, P. (1974): "Problems of Psychological Reality in Generative Phonology: A Critical Assessment", Reports from Uppsala University Department of Linguistics nr 4.
- MANDELBROT, B. (1954): "Structure Formelle des Langues et Communication", Word 10, pp. 1-27.
- MILLER, G.A. (1956): "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information", Psychological Review 63, pp. 81-97.
- OHALA, J.J. (1979): "The Contribution of Acoustic Phonetics to Phonology", to be published in LINDBLOM, B. and ÖHMAN, S. (eds.): Frontiers of Speech Communication Research, London: Academic Press.
- PIKE, K.L. (1943): Phonetics, Ann Arbor: The University of Michigan Press.
- STAMPE, D. (1972): "On the Natural History of Diphthongs", Papers from the 8th Regional Meeting, Chicago Linguistic Society, pp. 578-590.
- STEVENS, K.N. (1972): "The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data", in DAVID, E.E. and DENES, P.B. (eds.): Human Communication: A Unified View, New York: McGraw Hill.

UNIVERSALS OF LABIAL VELARS AND DE SAUSSURE'S CHESS ANALOGY

John J. Ohala, Phonology Laboratory, Department of Linguistics,
University of California, Berkeley, California, U.S.A.

In the Cours de linguistique generale de Saussure compares language to a chess game and the units of the linguistic code to the individual chess pieces. He remarks that

If I use ivory chessmen instead of wooden ones, the change has no effect on the system ... The respective value of the pieces depends on their position on the chessboard just as each linguistic term derives its value from its opposition to all the other terms ... [A single] move has a repercussion on the whole system [1916 (1966:22; 88-89)].

The choice of the chess analogy was a brilliant piece of exposition. Justifiably, it is frequently cited, especially by teachers in linguistic courses, and has become one of the favorite images of the structuralist basis of linguistic analysis.

The structuralist approach in phonology means analyzing a given problem by taking the whole phonological system into consideration, e.g. all the phonemic oppositions, especially those which are symmetrical or asymmetrical, the functional load of the sounds involved, etc. It focuses, therefore, on system-internal relations between the 'pieces', i.e., the speech sounds. For example, the structuralist account of the introduction of [ʒ] into English would point out that it filled what was up to that time a gap in the English fricative system:

f	θ	s	ʃ
v	ð	z	ʒ

Generative phonology, a recent offshoot of structural linguistics, also focuses on system-internal relations between the 'pieces' although in this case the pieces are the rules of grammar and the entities which make up the lexicon.

In fact, almost any post-Saussurean "school" of phonology one might cite, e.g., the Prague school, glossematics, functional phonology, natural generative phonology, etc. -- all have adopted the structuralist method of looking within the system for the solution to their problems. Occasional explorations outside the system -- into anatomy, physiology, physics, psychology, etc. have never been

pursued seriously or intently.¹

I would maintain that this emphasis on system-internal relations in phonology is counter-productive. This point is especially evident when we examine and seek explanations for phonological universals. We frequently find speech sounds behaving in very similar ways across languages even though those languages exhibit remarkably varied structure. The phonological behavior of labial velars, i.e., [u, w, m, \widehat{kp} , \widehat{gb}] etc. illustrates this rather dramatically.²

It has been claimed by generative phonologists that labial velars, although possessing two more or less equal constrictions, labial and velar, nevertheless, must be represented at the underlying (lexical) level as having only one primary articulation -- the other constriction being relegated to a secondary articulation (Chomsky and Halle 1968, Anderson 1976). The phonological behavior of a segment is supposedly a function of its underlying representation, not its surface phonetic character. Thus Anderson, in reviewing a number of West African languages, argues that Temne which has a /k/ but no /g/, must classify its / \widehat{gb} / as a velar, filling the gap in the voiced velar stop position. Similarly he argues that since Nkonya has both /k/ and /k^w/, the second sound thus preempting the classification: 'primary articulation: velar; secondary articulation: labial', the sound / \widehat{kp} / in that language must be primarily a labial with a secondary velar articulation. Efik, he notes, not only has a /k/ vs. /k^w/ contrast but also lacks a /p/, so it has two reasons for classifying its / \widehat{kp} / as a labial.

One of the problems with such structural or functional accounts of phonological facts is that they attach undue significance to sound patterns which may commonly arise due to chance or at least due to factors unrelated to the particular phenomena under investigation. Attention to phonological universals would be some insurance against this problem. As it happens, /p/ and /g/ are often missing from languages' stop inventories (Gamkrelidze 1975, Sherman

(1) Notable exceptions, however, are the fields of sociology, cultural history, and anthropology, which have been pursued seriously by many phonologists with structuralist orientation.

(2) The research on labial velars was done in collaboration with James Lorentz and published in Ohala and Lorentz (1977). Limitations of space prevent extensive documentation of the sound patterns discussed; however, the article cited may be consulted for numerous cross-linguistic examples.

1975). Moreover, there are many languages in West Africa that have / \widehat{kp} / and/or / \widehat{gb} / (Ladefoged 1964). Why therefore assume there is a special relationship between these two patterns in those few languages in which they both appear? A very preliminary statistical analysis of the co-occurrence of these patterns by Ohala and Lorentz (1977) found no disproportionate incidence of labial velar stops in languages which also have gaps in their stop inventory.

The most serious problem with such structuralist arguments, however, is that they often as not conflict with the evidence one can obtain from phonological alternations, including allophonic variation:

- 1) In spite of the double motivation mentioned above for assigning the Efik / \widehat{kp} / to the labial slot (as well as an additional reason, cited by Welmers 1973, namely, that / \widehat{kp} / sometimes is realized as the allophone [p]), Cook (1969) reports that a nasal assimilating to it sometimes appears as the velar nasal [ŋ].
- 2) According to Bearth (1971:18), Toura has both /k/ vs. /k^w/ and /g/ vs. /g^w/ contrasts, which, following the logic presented above, would force us to characterize / \widehat{kp} / and / \widehat{gb} / as labials. Nevertheless, these latter two sounds can be realized as [ŋ \widehat{kp}] and [ŋ \widehat{gb}], respectively, before nasal vowels.

Maybe one could still salvage the practice of looking only to system-internal relations in phonological analysis by abandoning the 'fill-the-gap' criteria and relying more heavily on how segments pattern in phonological rules. Unfortunately this escape route is not open either because labial velars can pattern in seemingly inconsistent ways in phonological rules.

- 3) The Yoruba labial velar glide /w/ (along with the labial velar stops / \widehat{kp} / and / \widehat{gb} /) patterns with the labials /b, f, m/ in that it causes the merger of following /ä/ with /ɔ̃/; nevertheless, the nasal assimilating to /w/ shows up as the velar [ŋ] (Ward 1952).
- 4) In Kuwaa (Belleh), word initial /w/ is occasionally realized as [ŋ^w], i.e. a labialized velar nasal, but may become labial [v] before unrounded vowels (Thompson 1976).
- 5) In Tenango Otomi /h/ becomes labial fricative [ɸ] before /w/ but /n/ assimilating to /w/ appears as [ŋ] (Blight and Pike 1976).

Additional such cases are not difficult to find (Ohala and Lorentz 1977).

The seeming confusion of these patterns is cleared up when system-external evidence is obtained, viz., data on phonological universals and the physical phonetic causes of the universals. I offer the following four statements of universal tendencies to account for the observed data:

A. When affecting the quality of adjacent vowels, labial velars behave primarily as labials. (Specifically, they cause vowels to shift in the general direction of [u].)

In addition to the evidence in 3, above, there is that from Tigre where, due to assimilatory action, certain short vowels are more back in the environment of labials, especially /w/ (Palmer 1962).

The phonetic basis of this pattern is the fact that labial velars achieve very low 1st and 2nd formant frequencies -- even lower than those of plain labials in most cases (Ladefoged 1964, Lehiste 1964) -- and thus are acoustically unlike sounds at any other than the labial place of articulation. This fact is itself capable of being explained by reference to acoustic phonetic theory (see Ohala and Lorentz).

B. When assimilating to adjacent vowels, it is the labial velar's labial place of articulation that remains unchanged; the place of the lingual constriction may shift or disappear under the influence of the vowel's lingual configuration.

Besides the evidence in 4, above, there is in addition the pattern from Dagbani in which the phonemes /k̄p̄, ḡb̄, ŋ̄m̄/ have the allophones [f̄p̄, d̄b̄, n̄m̄], respectively, before front vowels and the palatal glide /j/ (Wilson and Bendor-Samuel 1969).

There is no mystery about the causes of this tendency. Of the two constrictions of labial velars, only the lingual constriction is free to (partially) assimilate its place of articulation to that of adjacent vowels. The shift of the lingual constriction in such a case is exactly comparable to its shift in other velar consonants, e.g., [k, g, ŋ, x], whose lingual constriction -- as is well known -- is also influenced by neighboring vowels. The labial constriction, for obvious anatomical reasons, is not likely to shift its place of articulation via assimilation to that of the lingual constriction of adjacent segments.

C. When becoming a fricative or determining the place of articulation of adjacent fricatives by assimilation, [w] shows itself primarily as a labial.

In addition to the evidence in 5 (and possibly in 1) above, there are supporting statements such as the following by Heffner (1964: 160):

The fricative noises produced by the articulation of [French] [w] are slight, but such as they are, they come rather from the labial than from the velar constriction. Assuming that there are both labial and velar sources of fricative noise during these sounds, there are a number of possible phonetic reasons why the labial noise source should predominate. The most important is probably the fact that the configuration of the vocal tract anterior to the velar noise source (the airspace and the small labial constriction) constitute a low-pass filter that effectively attenuates the predominantly high frequency noise produced at the back constriction. The noise source at the labial constriction, of course, suffers no high frequency attenuation.

D. When becoming a nasal or determining the place of articulation of adjacent nasals by assimilation, labial velars behave primarily like velars.

Alongside the evidence in 1 through 5, above, there are many cases such as the dialectal variants for the word for "child" in two Melanesian languages: in Sa'a it is /mwela/ (which is more representative of the original form) but in Kwara 'Ae it is /ŋela/ (Ivens 1931).

The explanation for this pattern requires reference to the vocal tract configurations for the nasal consonants [m, n, ŋ] and [w̄] (to pick a common labial velar), which are represented schematically in Figure 1.

Essential to the acoustic characteristics of nasals are the pharyngeal-nasal airway and the oral cavity branching off from it. 'Oral cavity' here refers to that air space extending from the pharynx to the point of constriction. The oral configuration

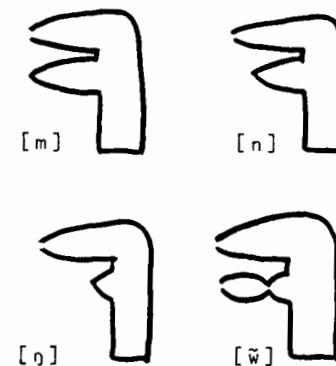


Figure 1.

anterior to the point of the rearmost constriction has no effect. It can be seen, therefore, that the acoustically relevant configuration of [w̃] is essentially similar to that of [ŋ].

It would seem from these data that the behavior of speech sounds is better understood by reference to system-external factors than system-internal factors. These are not isolated examples. A more appropriate analogy to offer as an image of language would be the game of football (American-style football). At any given time during a football game when the ball is in play, it is still the case, as in chess, that there is "significance" to the game in the special arrangement of the players, e.g. it is advantageous to the side possessing the ball to have an eligible receiver downfield. However, of more importance to the outcome of the game is the inherent ability of the individual players. It may not matter in chess whether one substitutes an ivory chess piece for a wooden one, but does matter in football if one substitutes a 50 kg tackle for one weighing 100 kg.

Conclusion

Observations of universal phonological tendencies -- for example, those found for labial velars, as in the present paper -- force us to the conclusion that the inherent physical constitution of speech sounds, i.e., how they are made and how they sound, have as much or more importance than system-internal relations, in determining the behavior of speech sounds. The emphasis most schools of phonology put on the study of system-internal factors is therefore a mistake.

Acknowledgment

The research was supported in part by the National Science Foundation.

References

- Anderson, S.R. (1976): "On the description of multiply-articulated consonants", JPh 4, 17-27.
- Bearth, T. (1971): "L'Énoncé Toura (Côte d'Ivoire)", Summer Institute of Linguistics, Publications in Linguistics and Related Fields, No. 30, Norman: S.I.L.
- Blight, R.C. and E.V. Pike (1976): "The phonology of Tenango Otomi", IJAL 42, 51-57.
- Chomsky, N. and M. Halle (1968): The sound pattern of English, New York: Harper and Row.
- Gamkredlidze, T.V. (1975): "On the correlation of stops and fricatives in a phonological system", Lingua 35, 231-261.
- Heffner, R-M.S. (1964): General phonetics, Madison: Univ. Wisconsin Press.
- Ivens, W.G. (1931): "A grammar of the language of Kwara 'Ae, North Mala, Solomon Islands", Bulletin of the School of Oriental and African Studies 6, 679-700.
- Ladefoged, P. (1964): A phonetic study of West African languages, Cambridge Univ. Press.
- Lehiste, I. (1964): "Acoustical characteristics of selected English consonants", IJAL, Publication 34.
- Ohala, J. and J. Lorentz (1977): "The story of [w]: an exercise in the phonetic explanation for sound patterns", Berkeley Ling. Soc., Proceedings 3, 577-799. Reprinted in: Report of the Phonology Laboratory (Berkeley), 1978, 2, 133-155.
- Palmer, F.R. (1962): The morphology of the Tigre noun, London: Oxford Univ. Press.
- de Saussure, F. (1966): Course in general linguistics. '[Transl. of original 1916 French ed.]', New York: McGraw-Hill.
- Sherman, D. (1975): "Stops and fricative systems: a discussion of paradigmatic gaps and the question of language sampling", Working Papers on Language Universals (Stanford) 17, 1-31.
- Thompson, R.B. (1976): A phonology of Kuwaa (Belleh), M.A. thesis, San José State Univ.
- Ward, I.C. (1952): An introduction to the Yoruba language, Cambridge: Heffer.
- Welmers, W.E. (1973): African language structures, Berkeley: Univ. of California Press.
- Wilson, W.A.A. and J.T. Bendor-Samuel (1969): "The phonology of the nominal in Dagbani", Linguistics 52, 56-82.

UNIVERSALS AND PHONETIC HIERARCHY

Kenneth L. Pike, Summer Institute of Linguistics
7500 W. Camp Wisdom Road, Dallas, Texas 75211

1. The Presumed Theoretical Basis for Some Past Avoidance of Syllable and Stress Group

In the mid 50's I cited evidence (Pike 1955, 66-68, amplified in 1967, 409-23) that on the American scene--and sometimes elsewhere also--the syllable had been often ignored, or denied theoretical status, or occasionally used without theoretical justification to support statements about the distribution of phonemes. Specifically, we might add that in Bloch and Trager (1942), in the chapter on phonetics, there is no section for the syllable (although there is one page--28--on 'Syllabic consonants' in which the syllable concept is used as background to the analysis). Similarly in the section on 'Semivowels' (22) syllabics are related to sonority, and syllables to syllabic sounds, with vowels treated as sometimes--but not always--syllabic. Later, in the chapter on phonemics, in the subsection on 'Vowels' (50) the syllable is used as a basis for discussing the distribution of simple vowels with strong stress, and related matters. But nowhere does the syllable as such get specific treatment in its own right as a basic unit of the system.

The reason: The underlying theoretical construct moved from the phoneme level to the morpheme level and on up to syntax, without the concept of syllable entering in as a level. They felt that a morpheme could be adequately described, in so far as its physical components were concerned, as made up of a sequence of phonemes. But if they had brought in the syllable as a basic unit of the system, there would have been much greater difficulty in justifying their descriptions, since oftentimes in ordinary speech a morpheme may be found which is either less than a syllable or more than a syllable, so that this leads to borders between units of the lexicon which would have been skewed with reference to those of the phonology. Thus the plural allomorph -s, is a single nonsyllabic consonant; but cups is a single syllable of two morphemes; and the morpheme ticket is a single morpheme of two syllables. Therefore, there could have been no direct mapping of (phonological) part to (morphological) whole if the syllable

had been treated as a unit in its own right.

2. A Theoretical Basis for Allowing Syllable, Stress Group, and Higher Level Phonological Units

In order to allow syllable, stress group, and even higher level units into our practical description, as units appropriate to that description, we need to have a theory of hierarchy which is multiple. Instead of a single hierarchy from phoneme to morpheme to syntactic unit, we need a hierarchy of phonology in its own right (from phoneme to syllable to stress group to phonological paragraph to phonological discourse--or something related to such a construct), and we need a hierarchy of grammatical units (from class of morphemes, to class of words, to class of clauses, class of sentences, class of paragraphs, and ultimately up to discourse classes), and in addition we need a referential hierarchy (of participants, episodes, and events as spoken about). The grammatical hierarchy (the telling order) may be distinct from the referential hierarchy (the happening--or logical--order, see Pike and Pike 1977, 363-410). Such a set of hierarchies in the theory allows us to have the syllable present in our description, and to draw upon it without apology (and without "boot-legging" it into the description).

This approach also allows us to specify openly some universals (e.g. no language is made up wholly of vowels) even though in some of them we may not find syllables composed of vowel plus following consonant. On the other hand, it does not insist that every possible level be present in every possible language. It insists, rather, that there be some hierarchical structure above the phoneme, without demanding that the syllable as such must inevitably be an emic unit. My personal suspicion would be that the syllable should be such a universal emic unit. But we have to leave room to the contrary, unless or until someone shows that the material on Bella Coola by Newman (in which the syllable is not treated as relevant) is not a satisfactory description (for preliminary discussion see Pike 1967, 420-21). Similarly, the work of Kuipers on Kabardian would have to be shown as better re-analyzed from a syllabic point of view (possibly by showing that he, like Bloch and Trager, relied on syllable without making adequate place for it in his theoretical system, for references see Pike 1967, 423).

The hierarchical approach also opens the door to the handling

of phonological markers of units much larger than a sentence (for example, the phonological paragraph). And in between the stress group and the phonological paragraph there may be emic sequences of stress groups (sequences of intonation contours) which have some overriding rising or falling general drift (or "tangent") within clause or sentence (see Bolinger 1970). And, above this, one may expect to find phonological units which signal the audience that a speaker is getting under way, or is finishing, or is changing focus. It should also be noted that there is strong evidence (overwhelmingly persuasive to me) that the kind of dynamic crescendo (or decrescendo) pattern of stress groups may in some languages be sharply contrastive within the styles of a single system. A greeting style, or a chanting style, or narrative pattern may, for example, affect these shapes; see Pike 1957, for example, for abdominal pulse types in inland Peru. A mark for juncture, plus a stress mark, is far from adequate to represent these contrasts; there must be both contrastive peaks and contrastive slopes leading down toward an end point (not just a stress mark followed by a final fade into some kind of "juncture").

3. Pairing in the Phonological and Grammatical Hierarchies

But the phonological hierarchy is not as simple as it sounds. There is no one direct sequence from phoneme to phonological discourse which meets some of the requirements for describing certain kinds of data which have an impact on us. Specifically, one of the most interesting developments--from my point of view--is that of Tench (1976). Tench was going beyond preliminary work on paired levels of the grammatical hierarchy (see now Pike and Pike 1977, 21-28) in which there was a sharp difference between units which are isolatable in the sense that (like an independent clause or an independent sentence) they could come at the beginning of a monologue, or at the beginning of a conversation after the greeting forms; and these would be in sharp contrast to responses to utterance, when the responses might sometimes be single words or phrases. This had led Pike and Pike to the setting up a difference between independent clause or sentence (as serving the function of serving as a proposition) versus word or phrase serving as a term. Tench showed a parallelism of these facts with the phonology, in which the syllable is the minimum independent item analogous to clause, while the rhythm group is the analogue of the

independent complex sentence. Similarly, he showed that the single phoneme (e.g. a single consonant), is analogous to a word (which is not isolatable in the same way) and that the consonant cluster would be the expanded version of that item, analogous to the phrase.

4. On Digital Versus Analogic Elements

More work needs to be done, also, to check out possibilities of digital versus analogic phonological structures. The digital ones (as pointed out by Martin and Pike 1975) are contrastive (either-or) units, the analogical elements have gradient (less to more) relation to the referent. My expectation would be that in every language we would find some analogic features of intonation and voice quality, in which length, loudness, rate, pause, decrescendo, crescendo (or features such as intensity, key, tenseness of vocal chords, breathiness), might be relevant in a gradient way, emphasizing the involvement of the speaker to a greater or lesser degree, or associated analogically with excitement or intensity of attitude.

But we would have to avoid assuming that such features were automatically to be found as digital in every language. For example, in Comanche (U.S.A.) no digital (contrastive, "segmentally phonemic") intonation elements have been found (Smalley 1953, 297).

The English-speaking actor on the stage, furthermore, is likely to make much greater use of the analogic types (change of key, for example), than is the ordinary person in a non-emotional setting. Yet our study of the systemic nature of contrastive quality is still in a very primitive state. It is astonishing that changes in voice quality seem to me to be empirically universal, but that a systemic handling of these materials is still only vaguely present with us. A "list" of voice qualities is far from satisfactory in handling the n-dimensional space which seems to be implicit in the possibility of simultaneous voice qualities, overlapping with pitch of various kinds, and interrupting (noncoterminal) units of the segmental phonological hierarchy from phoneme through syllable on up to phonological discourse. A vast amount of work seems to me to be awaiting us on the theoretical and empirical facets of these matters.

A final note: I am aware that there are difficulties in

finding physical correlates for perceived syllables. But I am convinced that any failures to do so in the past should not prevent us from continued search for something which is so obviously present in field work--since I cannot believe that a characteristic so universal can have no relation to some concomitant physical reality (no matter how complex the relation may prove to be).

References

- Bloch, Bernard, and George L. Trager (1942): Outline of Linguistic Analysis. Baltimore: Linguistic Society of America.
- Bolinger, Dwight (1970): "Relative Height", in Prosodic Feature Analysis, Pierre R. Léon, Georges Faure, and André Rigault (eds.), 109-25. (Reprinted in Intonation, Bolinger, ed., 137-53, Harmondsworth: Penguin.)
- Martin, Howard R., and Kenneth L. Pike (1975): "Analysis of the Vocal Performance of a Poem: a Classification of Intonational Features", Lg. and Style 7, 209-18.
- Pike, Kenneth L. (1957): "Abdominal Pulse Types in Some Peruvian Languages", Lg. 33, 30-35.
- ____ (1967): Language in Relation to a Unified Theory of the Structure of Human Behavior. Second edition. The Hague: Mouton. (First edition Vols. 1-3, 1954, 1955, 1960.)
- ____, and Evelyn G. Pike (1977): Grammatical Analysis. Summer Inst. of Linguistics Publ. in Ling. 53.
- Smalley, William A. (1953): "Phonemic Rhythm in Comanche", IJAL 19, 297-301.
- Tench, Paul (1976): "Double Ranks in a Phonological Hierarchy", J. of Ling. 12.1-20.

BASES FOR PHONETIC UNIVERSALS IN THE PROPERTIES OF THE SPEECH
PRODUCTION AND PERCEPTION SYSTEMS

Kenneth N. Stevens, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, U.S.A.

This paper discusses how the properties of the human articulatory and perceptual systems play a role in determining certain phonetic universals. In particular, our concern is with the inventory of phonetic segments that are found in language, and the way in which these segments are organized into a set of natural classes. We shall review how the articulatory and the perceptual systems place certain constraints on the classes of sounds that are used universally in language. The classificatory features that play a role in the phonological rules in language are determined by these natural classes that are based on observation of the capabilities of the articulatory and perceptual mechanisms.

Articulatory evidence for natural classes of speech sounds.

The actualization of a given speech sound in context requires a complex sequence of articulatory activity. The articulatory structures must be maneuvered from positions or states appropriate to one sound to states corresponding to the next sound to be produced. We shall follow the traditional method, used by phoneticians for years, of specifying a phonetic segment in terms of a set of goals or target states that the articulators are to achieve or that are intended by the speaker rather than in terms of the movements between these targets. The hypothesis is that these target configurations or states, if appropriately specified for a given sound, are much less dependent on the phonetic context than are the articulatory movements or muscle contractions necessary to produce the sound in context. Thus the articulatory descriptions are static, in the sense that they describe stationary states or configurations. While the production of some sounds or sound sequences may involve movement, this movement is always from one target state to another.

How are these articulatory target states to be described and how does this description lead to a specification of natural classes of speech sounds? Examination of lateral radiographs

gives us one view of the articulatory target states in terms of the positions of the various articulatory structures that are visible on the midline. This kind of evidence has traditionally been used in phonetics to describe articulatory targets in terms of place of articulation identified along the length of the vocal tract. Another way of describing articulatory configurations examines the pattern of contact that occurs between structures such as the tongue and palate. This pattern is presumably registered in the talker's speech control system through the responses of receptors located on the surfaces of the structures (Stevens and Perkell, 1977). Still another aspect of the target state is the physical properties of the surfaces of the structures, particularly the vocal folds and the tongue. These properties have an influence on the manner in which the airflow from the lungs is controlled and on the way in which the articulatory structures are forced against one another. Which of these ways (or combinations of ways) of describing articulatory states is most salient for grouping speech sounds into natural classes is a question about which we can only speculate at present.

We consider now several lists of phonetic segments. For all of the items on a given list, some aspect of the articulation is achieving the same state, as defined in at least one of the ways listed above. We suggest, then, that these items can be candidates for forming a natural class of phonetic segments.

[m n ŋ ã õ ...] These items are all produced by creating velopharyngeal opening, usually by placing the velum in a lowered position. From the point of view of the speaker, an indication that the velum is lowered comes from several possible sources: (1) the muscles used to lower the velum have been contracted; (2) the lowered state of the velum is sensed through receptors that signal the position of the velum or its contact with other structures; (3) there is airflow through the velopharyngeal opening and possibly acoustic energy in the nasal cavity that is sensed and registered in some way.

[k g ŋ i u ...] These sounds are all produced by placing the tongue body in a raised position within the oral cavity. More specifically, the common articulatory activity for the sounds can

be described in one of two ways: (1) there is contraction of a common muscle or group of muscles to produce the raised tongue body, or (2) there is a common pattern of activity in particular groups of sensory receptors in the tongue musculature or on the dorsal surfaces of the tongue as these surfaces make contact with other structures, particularly the hard palate (Stevens and Perkell 1977).

[p t k č f θ s š á í ú ...] For this group of sounds, it is hypothesized that the common articulatory attribute is a stiffening of the surfaces of the vocal folds (Halle and Stevens, 1971). The articulatory state that characterizes each member of this class can be described either as contraction of a particular laryngeal muscle or group of muscles or as the stiffened state of the vocal fold surfaces, independently of the muscle activity used to produce that state.

[p t k č b d g ĵ m n ŋ ...] The sounds in this group are all produced by forming a complete closure of the vocal tract at some point along its length. The articulatory description for this group of segments cannot be specified in terms of the contraction of particular muscles, since different muscles are clearly involved depending on where in the vocal tract the constriction is made. Rather, it is assumed that an instruction to form a complete closure is a basic component of articulatory control which, when coupled with a further instruction indicating which articulator is to be activated, effects the proper consonantal constriction. It is possible also that the sensory consequences of forming a complete closure are registered in some unique manner independently of the location of the closure in the vocal tract.

[p b f v m ...] The segments on this list have the common articulatory attribute that they are produced with a constriction at the lips. Thus a particular set of muscles - those making a lip closure - is involved in the generation of all of these sounds. The lower lip comes in contact with either the upper lip or the upper incisors, and this gesture leads to a unique pattern of excitation of sensory units in the lower lip.

[t d n θ ó s z š ž ʎ r ...] These phonetic segments are all actualized by raising the tongue blade to make contact with some

part of the maxilla. The exact region of contact or the force of contact may vary from one sound to another in the set, but the common gesture is that of raising the tongue blade, presumably through contraction of certain intrinsic tongue muscles. There is a unique sensory consequence of this raised pattern of the tongue blade: the edges of the superior portion of the tongue come in contact with fixed surfaces of the hard palate or teeth, presumably leading to a special response of tactile receptors on these surfaces of the blade.

The six lists of segments given above are examples of a longer inventory of lists of segments that could be generated. Furthermore, there is no attempt to make each list exhaustive; additional items could be appended to the lists. These examples serve to indicate, however, that natural classes of speech sounds can be constructed through examination of the articulatory target configuration or states. In giving these examples, we have shown a certain amount of ambivalence as to how the common articulatory attributes for the items on a list should be specified. Until we know more about how motor systems operate, and, in particular, how the speech-production systems operate, the question of how best to characterize natural classes of speech sounds in terms of articulatory attributes must remain open.

Acoustic and psychoacoustic evidence for natural classes

Acoustic analysis of speech shows that there are groups of speech sounds that share common acoustic properties. If it is assumed that the auditory system responds in some unique way to sounds with a common acoustic property, then this unique response provides the listener with a means for organizing speech sounds into natural classes based on their acoustic properties. As examples, we shall consider several lists of speech sounds, and we shall show that for the items in any one of these lists there is a common distinctive acoustic property. The basis for these classifications is derived largely from the work of Fant (1960), Jakobson, Fant and Halle (1963), and others.

[m n ŋ] For the items on this list, there is a rather steady nasal murmur persisting for several tens of milliseconds, with an amplitude just a few dB below that of the adjacent vowel. The unique acoustic attribute of this nasal murmur is a strong

spectral peak at low frequencies and a relatively uniform distribution of weaker spectral peaks at higher frequencies, with these peaks tending to be rather broad (Fujimura, 1962).

[t d n s z ʒ ʒ̃ ʒ̃̃] For these consonants, the spectrum sampled at or near the consonantal release (in a consonant-vowel syllable) shows a diffuse spread of energy across the frequency range, but with greater spectral energy at high frequencies (Fant, 1960; Zue, 1976; Stevens and Blumstein, 1978).

[k g ŋ] The spectrum at the consonantal release for these sounds has a single prominent peak in the midfrequency range (Fant, 1960; Zue, 1976, Stevens and Blumstein, 1978).

[i ɪ u u] The vowels in this list all have a relatively low first formant.

[ã ü ɪ̃] These nasalized vowels have a spectrum in which the lowest peak, corresponding to the first formant region for a nonnasal vowel, is split or broadened to cover a wider frequency range than that for a nonnasal vowel.

[p t k ʧ b d g ʝ m n ŋ] The items in this list all show an abrupt onset of spectral energy over much of the frequency range when the consonant is released into the following vowel. The rise in amplitude in any one frequency region occurs in a time interval of just a few milliseconds. A sound with an abrupt onset has been shown to produce a distinctive response in a listener (Cutting and Rosner, 1974).

[ɟ ɟ̃ ɟ̃̃] These vowels all have a fundamental frequency (F_0) that is high in comparison with the average F_0 for the particular speaker and the particular position of the vowel within an utterance.

[p t k ʧ f θ s ʃ] The common acoustic characteristic of the sounds in this list is the absence of low-frequency periodicity in the sound in the vicinity of the consonantal closure interval.

As in the case of the lists based on articulatory attributes, the above lists are examples of a longer inventory of lists such that the items in each list have a common acoustic property to which the auditory system is assumed to respond in a unique way. Given our present rudimentary knowledge of the response of the auditory system to complex sounds, we have only

been able to speculate on the kinds of acoustic properties that qualify for defining groups of speech sounds.

The classificatory features

Examination of the two sets of lists - these based on common articulatory attributes and those based on common acoustic attributes - reveals that there is much overlap in the two sets. This overlap is not surprising, since on the basis of acoustical theory it is not unexpected that sounds produced with common aspects of the articulatory configuration should also have similar acoustic characteristics.

Another way to organize speech sounds into natural classes is to examine the phonological rules of language, and to observe the various groups of segments that are operated on by these rules or that determine the environments in which the rules operate. The grouping of segments according to this criterion leads to a description of segments in terms of bundles of classificatory or distinctive features. These classificatory features also show a great deal of overlap with the groupings based on articulatory and acoustical considerations.

We would like to propose a rather simple condition on the definition of a classificatory feature: a set of speech sounds shares the same classificatory feature if the sounds share a common articulatory attribute and a common acoustic or perceptual attribute. That is, the sounds in a given class should give rise to response patterns that have a common property in the auditory system of the listener and the speaker, and, in addition, the production of the sounds should have common attributes in the speech-generating mechanism of the speaker, such as common patterns of orosensory response.

A consequence of this definition is that vowels and consonants will tend not to share the same features. Thus, for example, nasal vowels and nasal consonants would not have the same feature, although it might be desirable to mark in some manner the fact that they share an articulatory property. The strong definition of a classificatory feature would not capture in terms of feature specifications the fact, for example, that vowels preceding nasal consonants tend to be nasalized (or in fact that nasalization of the vowel often is accompanied by

elimination of the consonant), or the fact that the pitch of vowels following voiceless consonants tends to be raised. These kinds of modifications are, in a sense, simply mechanical consequences relating to the coarticulation that is a nature consequence of the juxtaposition of two segments.

The classificatory features defined in the way we have proposed would, however, specify major classes of segments that play a role in the phonological rules of language. These features would owe their existence, so to speak, both to the property-generating characteristics of the speech production system and to the property-detecting characteristics of the speech perception system.

References

- Cutting, J. and B. Rosner (1974): "Categories and boundaries in speech and music", Perc. Psych. 16, 564-570.
- Fant, G. (1960): Acoustic theory of speech production, The Hague: Mouton.
- Fujimura, O. (1962): "Analysis of nasal consonants", JASA 34, 1865-1875.
- Halle, M. and K.N. Stevens (1971): "A note on laryngeal features", Research Laboratory of Electronics Quarterly Progress Report No. 101, M.I.T., Cambridge, Massachusetts, 198-213.
- Jakobson, R., G. Fant, and M. Halle (1963): Preliminaries to speech analysis, Cambridge, Massachusetts: M.I.T. Press.
- Stevens K.N. and S.E. Blumstein (1978): "Invariant cues for place of articulation in stop consonants", JASA 64, 1358-1368.
- Stevens, K.N. and J.S. Perkell (1977): "Speech physiology and phonetic features", in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.), 323-341, Tokyo: University of Tokyo Press.
- Zue, V.W. (1976): Acoustic characteristics of stop consonants: A controlled study, Ph.D. thesis (unpublished), M.I.T., Cambridge, Massachusetts.

THE PSYCHOLOGICAL REALITY OF PHONOLOGICAL DESCRIPTIONS

Summary of Moderator's Introduction

Victoria A. Fromkin, University of California, Los Angeles,
California 90024, USA

A phonological description of a language will be a 'true' description to the extent that it is 'psychologically real'. A theory of phonology will be a 'true' theory to the extent that it permits the construction of psychologically real grammars. These assumptions are required of an empirically based phonological theory. What we seek then is evidence that will help decide whether a particular description is 'psychologically real'. There are no a priori principles which can be depended on. We do not know in advance whether, for example, the human mind can or does relate two levels of phonological representation--phonemic and phonetic--by ordered rules, nor do we know the extent to which the immature child's brain can draw highly abstract generalizations from a limited set of input stimuli. In fact, we have not progressed too far since 1887 when Fournie observed that "Speech is the only window through which the physiologist can view the cerebral life". Psychologists, neurologists, and linguists depend, to a great extent, on linguistic facts to determine the capabilities of the human mind. We have not found any direct ways, as yet, to observe what is "'in people's heads' (and) since we cannot look into people's heads directly we can only hypothesize what goes on there on the basis of indirect evidence" (Chafe, 1970). Even when we do look into people's heads directly, we cannot find in the physical brain matter, in the 10^8 neurons, or even in the neural organization of the cortex, the information we seek regarding the nature of the internalized grammars, the information which will tell us whether our theory, or which theory, of phonology is psychologically real or 'true'.

This symposium is concerned with the kinds of evidence which will help decide this question. While we all seem to agree on our aims (at least to the extent that we seek 'psychological real grammars') we are not necessarily in agreement as to what counts as evidence, how to weigh different kinds of evidence, or even what is meant by 'psychological reality'.

Cutler suggests a division between the proponents of a

'strong sense' as opposed to a 'weak sense' of psychological reality. The first group considers levels (e.g. phonemic representations) and processes (e.g. P-rules) to be psychologically real if a processing model includes stages isomorphic to levels and mental operations corresponding to the processes or rules. Linell also refers to this division. Cutler's paper presents speech error data to show that lexical stress and word formation rules are psychologically real in the weak sense, but not in the 'operational' or 'strong' sense. Linell also suggests that "rules must not be equated with behavioral processes...(since) conventional phonological rules state nothing but regular correspondences between idealized representations of the same or related pronunciations." In the fuller version of my paper I will discuss some evidence from speech errors which suggests that at least some rules and some levels are real in the strong sense of the term, but that this should not be a criterion for a theory of phonology.

Derwing's paper seems to support the 'strong' view. For example, he questions "what psychological sense can possibly be made...of a notion of 'rule ordering' which has no relation to real time" and further proposes that "if grammars relate in any way to psychological events or states (my emph.) then we need to interpret grammars psychologically." Grammars can, however, 'relate' to events or states without being identical or even isomorphic to them. And one can conceive of ordered relations, hierarchical for example, in a non-behavioral way and on a non-real-time basis. The alphabet may be represented in memory ordered from A to Z even for a brain damaged patient who cannot retrieve the letters in that order in real time. Cognitive psychologists concerned with lexical storage are providing evidence for intricately ordered classification systems based on ordered basic and primary levels of categorization in the levels of abstraction in a taxonomy (Rosch, 1978). Derwing also discusses aspects of the question which relate to the philosophy of science (as do Linell and Skousen), some points of which I will further discuss. But it is clear that whether a theory or a grammar is psychologically real must depend on empirical evidence rather than one's philosophical biases.

Bondarko's paper is neutral as to some of the controversies discussed in the other papers, positing three psychologically

real levels of phonology--production and perception of speech sounds, the phonemic level, and the level of word formation rules--as evidenced by perception experiments.

Campbell, Dressler, Gussman, and Skousen, are concerned with the importance of internal versus external evidence in the testing of linguistic hypotheses and the evaluation of theories. Internal evidence refers to facts drawn from the overall grammar, significant generalizations, simplicity factors, distributional criteria, morphemic alternations etc. External evidence refers to acquisition data, language disturbance, borrowing, orthography, speech and spelling errors, metrics, casual speech, language games, historical change, perception and production experiments etc. (Cf. Zwicky, 1975). Campbell and Skousen, and to a certain extent, Dressler, place major emphasis on external evidence. Campbell is very convincing in his demonstration of how language games in Finnish and Kekchi, for example, strongly support the reality of a vowel harmony rule and a vowel-epenthesis rule, respectively. He provides similar evidence in support of morpheme structure conditions as opposed to syllable structure rules. Skousen uses similar arguments. But Dressler shows that external evidence can be contradictory and Gussman provides some detailed illustrations supporting this. Interestingly, where Skousen posits external evidence from tongue slips to show the correctness of analyzing the affricates in English as non-sequential units, /č/ and /ǰ/, Gussman provides other external evidence, i.e. low level phonetic rules, which argue for the sequential analysis. Gussman points to the Fromkin (1971) data cited by Skousen to illustrate this contradiction. He also ties in the question of 'abstractness' with 'psychological reality' and correctly, I believe, shows that the question should not be how abstract is an analysis, but is it right or wrong. An important question to be discussed in the symposium, then, is what to do when different kinds of evidence are contradictory. It is also important for us to clarify how both internal and external evidence are to be used. If we find in Kekchi, for example, that an experiment on loan words supports morpheme structure conditions is this to be used only for the grammar of Kekchi or as evidence for the meta-theory of phonology? If speech error data argue for a rather abstract representation in some language, is this evidence that one can provide such abstract

representations in all languages? In other words, are we looking for evidence as to constraints on a general theory of phonology or for evidence concerning a grammar of a particular language?

Given the extent to which individual grammars may vary across speakers of one language, should we not seek constraints on the general theory which will permit us to construct the optimal, 'psychologically real' grammar for a language? The papers already cited reveal the problems we face. Data alone, and multiple-kinds of evidence alone will not provide all the answers. We need universal principles and a theoretical framework which in a principled fashion will help us constrain phonological descriptions to psychologically real ones. Skousen presents such a principle-- a principle of maximizing acoustic differences. Hale's paper is primarily concerned with just such questions and posits a 'principle of recoverability', with supporting evidence from Papago and Maori. What we need is more principles, supported by clear empirical evidence. For we can probably all agree that "However difficult it may be to find relevant evidence for or against a proposed theory, there can be no doubt whatsoever about the empirical nature of the problem" (Chomsky and Halle, 1968).

References

- Chafe, W. (1970): Meaning and the Structure of Language, University of Chicago Press, Chicago.
- Chomsky, N. and M. Halle (1968): The Sound Pattern of English, New York: Harper and Row.
- Fromkin, V.A. (1971): "The non-anomalous nature of anomalous utterances", Language 47, 27-52.
- Rosch, E. (1978): "Principles of categorization", in Cognition and Categorization, Eleanor Rosch and Barbara B. Lloyd (eds.), 27-48, Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Zwicky, A. (1975): "The strategy of generative phonology", in Phonologica 1972, Dressler and Mareš (eds.), 151-168.

ON THE PHONOLOGICAL OPERATIONS ENSURING SPEECH COMMUNICATION

L.V. Bondarko, Department of Phonetics, University of Leningrad, USSR

Conveying information through articulate speech presupposes the ability of the native speaker to analyse quickly and effectively heterogeneous sounds. This ability is developed by man because sound differences are used for discriminating meaningful units, i.e. words. Taking this function of speech sounds into consideration, we can understand why the native speaker does very well in the process of perception in spite of a number of variations of sound properties. From the linguistic point of view it can be assumed that there exist a number of levels ensuring optimum processing of sound signals. The first one consists in the ability of man to generate and perceive articulate sounds. Though this ability is universal by itself, it cannot be observed directly because it is realized on the basis of a certain concrete language. However, some of the phonetic universals (Greenberg, 1966) deduced on the basis of comparing various languages, can be also related to the peculiar properties of man's verbal behaviour. The second level is concerned with the system of phonemes in a given language. The native speaker disposes of the information of the system of phonemes which he acquires in the process of learning his native language. The main points of this information are as follows: the inventory of the phonemes in the language, the ways the distinctive features of the phonemes are realized, the rules of usage which include the probability of the occurrence of phonemes within the minimal meaningful unit - within a word.¹

The third level deals with the information of the rules about possible sound combinations in shaping the words. One can assume that the perception of the word is the recognition of its phonemic composition. Evidently a clear-cut differentiation of all the three levels is impossible, because practically they overlap to a great extent. But one may hope that the systematic research on the process of perception will enable the scientists to describe these levels in a more detailed way.

(1) It is possible that in a number of cases a morpheme may be treated as this minimal unit. This may take place in languages where phonemic alternations are regular and are governed by the existing rules, Russian being an example.

Let us consider some facts dealing with each of these levels which testify to the reality of the language consciousness of the speakers. The opposition of consonants with regard to "absence - presence of voice" is one of the most widespread (Zhivov, 1976). In fact, it can be connected not only with the function of the vocal cords alone, but also with properties like tenseness - laxness, delay in the onset of voice after the opening of the occlusion, the duration of the preceding vowel, and so on. One may assume that "absence - presence of voice" can be treated as a universal feature. For the native speaker of the Russian language, where the correlation "presence versus absence of voice" is one of the characteristic features, each consonant he hears must be described either as a voiceless or as a voiced one. But the consonants /c/, /č/, /x/ do not have voiced correlates, i.e., the opposition of voiceless consonants to voiced ones is not possible for them in the positions before vowels and consonants. Compare [tu'goj] - [du'goj], [sɪpɪtʃ] - [gɪbɪtʃ] and [tsex], [tʃaj], [xot], and so on. However, in accordance with the rules of alternations which are known to be regular in the Russian language, in the combination of words ending in the consonants /c/, /č/, /x/ ([ts, tʃ, x]) with words in which initial consonants are voiced obstruents, there appear voiced allophones of these voiceless consonants: [kan'nedz zɪ'mɪ], [zedz drɐv'va], [moɣ ga'ʃit], phonologically: /kan'éc z'i'mɪ/, /žeč drɐv'vá/, /moɣ gar'ít/.

The voiced character of these phonologically voiceless consonants can be treated in various ways from the linguistic point of view. We are especially interested in how the voiced character is treated by the Russian native speaker who is expected to discriminate between voiceless and voiced consonants and who does not have at his disposal the voiced correlates of phonemes which possess the same properties as /c/, /č/, /x/.

Russian subjects when presented with the consonants from phrases of the type /kan'éc z'i'mɪ/, /žeč drɐv'vá/, /moɣ gar'ít/, cut out from the magnetic tape, recognized these consonants as voiced ones; other properties of the consonants could be perceived incorrectly in this case. If the phonetic context is enlarged and the subjects are presented with combinations - 1: including the following consonant (CC), 2: including also the preceding vowel (VCC), 3: including the vowel in the succeeding syllable as well, - the

recognition of the consonants under consideration as voiced ones occurs less frequently, though in these cases the consonants /c/, /č/ and /x/ are not interpreted 100% correctly.

Figure 1 presents data on how separate properties of the consonants /c/, /č/ and /x/ are perceived if they are presented in various contexts, such as C, CC, VCC and VCCV. The influence of

the phonetic features proper increases with the narrowing of the phonetic context, although even if there is a complete phonetic context - the following consonant bringing about voicing, or vowels, ensuring as a rule good recognition of the neighbouring consonant - this is not sufficient for the recognition of such phonemes as /c/, /č/ or /x/. The sounds may be perceived as /c/, /č/ or /x/ only if the native speaker hears the whole phrase, i.e. if he makes use of both the phonetic and the semantic contexts (Bondarko, 1975). This means that the predominant influence of the first, universally phonetic level is removed only if both the second level including rules of alternations, and the third level concerned with the

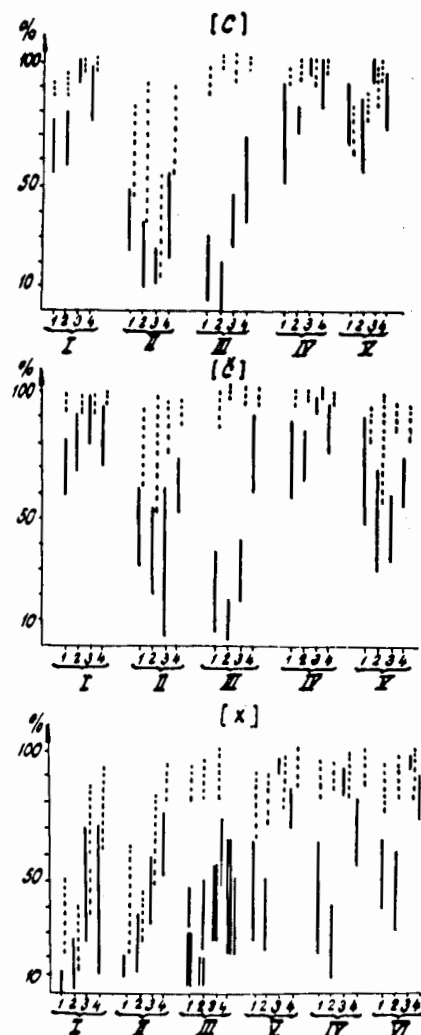


Figure 1

The perception of the properties of the voiced (—) and voiceless (---) allophones of the consonants /c/, /č/ and /x/. The phonetic context: (1) -C, (2) -CC, (3) -CCV, (4) -VCCV. Properties: I the active organ of speech II the manner of production III absence - presence of voice IV noise - sonorous character V hardness - softness VI vowel-consonant character

analysis of the phonemic composition of words can be made use of.

The second level of analysing speech, as has already been mentioned, includes information about the inventory of phonemes in the given language, the ways in which the distinctive features are realized, and the rules of usage. It is this level that ensures the transition from the phonetic variations of real sounds to economic phonological interpretations. Let us consider this level of perception using the examples concerning the perception of vowels by Russian native speakers.

It is known that the system of vowels in the Russian language is comparatively poor. There are three degrees of height and two series. Vowels of the back series (with the exception of the lowest vowel /a/ are necessarily rounded, whereas this connection does not exist in the case of the front vowels. The six vowels /a/, /o/, /u/, /e/, /i/, /i²/ are realized differently in the stream of speech, depending on their stressed or unstressed character, the quality of the neighbouring consonants, and so on.

As was shown in an experiment (Bondarko et al., 1966), the i-like transition, appearing in the vowel under the influence of the soft neighbouring consonant, serves as a useful indication which enables a person to differentiate hard and soft consonants. The i-like transition (phonetically pushing forward the vowel into the front zone) is perceived by all Russian native speakers as a cue of the consonant. Nevertheless, the phonetic property itself is realized in the vowel, and Russian native speakers discriminate a greater number of vowels than could have been expected on the basis of the inventory of vowel phonemes in the language.

We can assume that it is this peculiarity in the realization of the feature of softness in consonants that enables Russian speakers to describe vowels of the type [y], [ø], [œ] at a universal, phonetic level. These are integrated in the inventory of vowels in the same way as is done by speakers of those languages in which these vowels represent phonemes (Slepokurova, 1971). Things are different in the situation where vowels adjacent to nasal sounds are presented. In this phonetic position, Russian

vowels are considerably nasalized and it could be expected that Russian speakers would use such changes in vowels by analogy with those that are observed in the position with the neighbouring soft consonants. But in reality, the results are quite different.

In a special investigation (Belyakova, 1977) dealing with the perception of nasal vowels of the French language and nasalized Russian vowels by Russian and French subjects, it was shown that French people recognize nasal vowels of their own language much better than Russians do theirs, but that they are less sensitive in the perception of Russian nasalized vowels. They perceive Russian nasalized vowels as non-nasalized. A comparatively low degree of the recognition of the Russian vowels a and e by French listeners can be accounted for not by the influence of nasalisation but by the influence of the neighbouring soft consonant, which leads to the perception of this vowel as more front and less open, i.e. a as e, e as i. It is typical of Russians to make a lot of mistakes in the recognition of the nasalized vowels (Fig. 2).

Finally, it is on the third level, dealing with the rules of the formation of the sound shape of the word, that a phonological interpretation of sounds is given, which has no unique phonetic correlate. For example, the recognition of the unstressed vowel

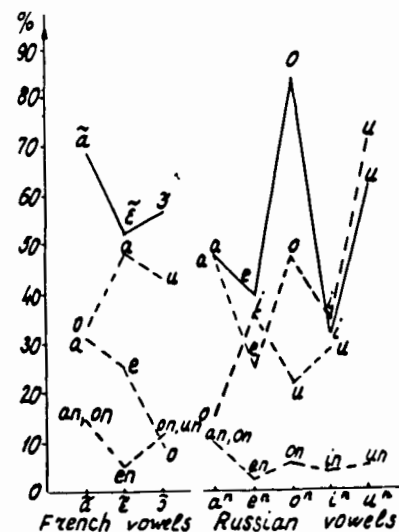


Figure 2

The perception of French nasal and Russian nasalized vowels. French listeners ——— Russian listeners ----- For all the subjects, various identifications of the vowels are shown: as the corresponding nasal vowel, as non-nasal but having different quality, and as a combination of a non-nasal vowel with a nasal consonant. Such identification is indicated in the figure by: an, en, and so on, even in those cases where the subjects wrote down the sounds an, am, etc.

(2) We do not consider here the question of the phonemic relevancy of the opposition of /i/ - /i²/, because it is widely discussed from the linguistic point of view, and, practically, because in the linguistic analysis it is not treated from the point of view of the phonology of the native speaker, for whom these are different vowels, and not on the lowest level alone.

in the words [sʌ¹rok], [dʌ¹ma] and so on, as /a/ is connected with the rules of reduction in the Russian word; the recognition of the voiced affricate as a voiceless one in the phrase "otec bolen" ([ʌ¹tʲedz¹boʎ¹n]) is connected with the rules of alternating voiceless and voiced consonants.

The recognition of morphologically loaded sounds or sound combinations represents a special case, particularly for such a language as Russian (Bondarko et al., 1966). In these cases the phonetic information about the sound is often insufficient, although the use of the rules of alternation and the use of semantic redundancy of the context enable the subject to correctly interpret the phonemic composition of the word (compare the realization of the phoneme /s/ in the combination "brosj ŝumetj" ([broʂ ʃu¹mʲetʲ]) with a considerable assimilation of /s/ to the following /ʂ/ and the realization of the phoneme /a/ in posttonic inflections after the soft consonant "njanja" ([¹nʲaŋʲ]), and so on.

All this proves that in oral communication, a person performs rather complicated operations the total of which can be called the phonology of the native speaker.

The reality of other purely linguistic phonological descriptions is proven by the extent to which this description is in accordance with these operations. The description of the phonology of the native speaker, based upon the description of different levels determining his verbal behaviour and upon the comparison with the linguistic phonology set up in linguistic descriptions, can be considered the main task in the experimental phonetic investigations dealing with speech perception.

References

- Belyakova, G.A. (1977): "The nasalization of vowels and its perception (on the basis of French and Russian languages)", Vestnik L.G.U., No. 8.
- Bondarko, L.V. (1969): "The syllable structure of speech and distinctive features of phonemes", Phonetica 20.
- Bondarko, L.V. (1975): "The phonemic description of the utterance - the condition and the result of understanding context", The minutes of the Fifth All-Union Congress of Psycholinguistics and the Theory of Communication, Moscow.
- Bondarko, L.V., L.A. Verbitskaya, L.R. Zinder and L.P. Pavlova (1966): "The sound units that can be distinguished in Russian speech", in The Mechanism of Speech-Production and Speech-Perception of Complex Sounds, Nauka, M.L.

Bondarko, L.V. and L.A. Verbitskaya (1975): "Factors underlying phonemic interpretation of phonetically non-defined sounds", in Auditory Analysis and Perception of Speech, London: Academic Press.

Greenberg, J. (1966): "Synchronic and diachronic universals in phonology", Language 42, No. 2.

Slepokurova, N.A. (1971): "On the position of the phonemic boundary between synthetic vowels", in The Analysis of Speech Signals by Man, Leningrad.

Zhivov, V.M. (1976): "The universals of syntagmatic functioning of the feature of voiceness", The Institute of the Russian Language of the Academy of Sciences of the U.S.S.R., A Study Group dealing with experimental and applied linguistics, preliminary publications, Issue 89, Moscow.

THE QUEST FOR PSYCHOLOGICAL REALITY: EXTERNAL EVIDENCE IN PHONOLOGY

Lyle Campbell, State University of New York at Albany, Albany, New York, U.S.A.

The principal goal of linguistic description is to account for a language in a way which reflects the competence of its speakers. This goal of achieving psychologically real descriptions (grammars) is reasonable. However, generative phonologists too frequently assumed that child language learners were somehow constrained to acquire the simplest possible grammar, and because the notation was designed to convert true generalizations into considerations of length, the simplest grammar the linguist could write was taken to be the psychologically real grammar. For this reason, arguments based on formal simplicity alone were considered sufficient to resolve many issues, e.g. abstractness, rule ordering, etc. However, formal simplicity (internal evidence) is not sufficient. Questions about psychological reality and the learnability of rules cannot be answered through considerations of surface patterns, distribution of allomorphs, combinatory properties of phonological elements within a linguistic system, and the like. The real question is, how can the linguist be certain that the rules he postulates to account for phonological patterns he perceives in his data correspond to the rules that speakers of the language establish? There are at present no successful formal criteria for determining from a given set of data what the speaker's rules may be.

A serious search for answers to this question must involve "external evidence", evidence not confined to surface-pattern regularities, but evidence which shows speakers behaving linguistically in ways where they must call upon their knowledge of the rules and underlying forms of their language in overt and revealing ways. The goal of this paper is to argue that external evidence should be given a stronger role in phonological investigation and to illustrate its potential.

Sources of external evidence that have been used with some success are: Metrics and verse (Kiparsky 1968, 1972; Zeps 1963, 1973), word games (Sherzer 1970; Campbell 1974, 1977) borrowing (Campbell 1974, 1976), experiments (Ohala 1974), speech errors

(Fromkin 1971, 1973), construction of orthographies, language change, etc. Here I will attempt to show the relevance of external evidence for validating aspects of individual grammars and for refining linguistic theory.

External evidence can demonstrate the psychological reality of certain rules. I will consider two examples, both involving word games (secret languages). The first is vowel harmony in Finnish. Since Finnish vowel harmony has many exceptions and complications, some have suspected that it may not be a psychologically real rule of Finnish grammar. In *kontti kieli* (or *kontin kieli*) "knapsack language", one of several Finnish word-game secret languages, the first consonant(s) and vowel of a word are replaced by ko (of *kontti*), and the material for which ko is substituted is placed before ntti (of *kontti*). Thus *veitsi* "knife" becomes *koitsi ventti*, *susi* "wolf" is *kosi suntti*. In this language game, vowel harmony adjusts the remaining vowels of the word to agree with ko (the harmonic series are back a, o, u, front ä [æ], ö [ø], y [y], and neutral i, e):

pysähtyköön "let him stop" becomes *kosähtukoon* *pyntti*
kylpylössä "in the baths" becomes *kopuloissa* *kyntti*
hänekö "him?" becomes *konkō* *häntti*

If vowel harmony were not a psychologically real rule of Finnish, speakers would not be able to adjust the vowels productively to agree with the back vowel when ko is substituted. (For full arguments, see Campbell 1977, in press.)

The second case is from Kekchi (a Mayan language of Guatemala). In *Jerigonza*, the Kekchi word game, one places p after each vowel followed by a copy of that vowel; for example *q'eqi^v?*, the name of the language, is *q'epeq^vipi?* in *jerigonza*. This game shows several Kekchi rules to be psychologically real, for example the rule of vowel-epenthesis before voiced labials ($\emptyset \rightarrow V_1/V_1C_{\{b^m\}}$) (examples: *kwiq'ib'a:nk* /*wiq'-b'ank*/ "to break it", *k'oxob'a:nk* /*k'ox-b'ank*/ "to seat something"). In normal speech, these forms never occur without the epenthetic vowel, but one may speak *jerigonza* optionally leaving out the epenthetic vowel: *kwipiq'b'apa:nk* or *kwipiq'ipib'apa:nk*, and *k'opoxb'apa:nk* or *k'opoxopob'apa:nk*. The rule of vowel epenthesis must be psychologically real; speakers must know the rule because they take it

into account in producing jerigonza forms -- they never leave out the wrong vowel, only the vowel which results from the rule of epenthesis.

These word games provide evidence for the reality of several other rules in these two languages, as well. Here, the external evidence helps resolve issues concerning the correct description of the individual grammars. External evidence has important implications for theoretical issues, also. I will present just one example, also from Kekchi.

Bilingual informants in Spanish and Kekchi were presented a list of loan words, some from Spanish into Kekchi and some from Kekchi into Spanish, and asked to judge whether the forms were borrowed, and if so, which they thought was the original language. Judgements were based on several parameters (cultural, semantic, and phonological). These parameters were determined by asking the informants why they thought particular loans to be Spanish or Kekchi in origin. Reasons volunteered by these informants involved, among other things, native views of morpheme structure in the two languages. For example, informants said pio:c̣ "pickaxe" (from Spanish piocha) and vilte:p "small chile" (from Spanish chiltepe), and similar forms, were from Spanish because Kekchi does not have those kinds of sounds together (vowel clusters in the first case, consonant clusters in the other). In actual fact, Kekchi does have vowel-vowel and consonant-consonant clusters, but only across morpheme boundaries (e.g. ke-ok' "get cold", ke- "cold" plus a verbal suffix), but never within a morpheme. This shows that these morpheme structure conditions of Kekchi are psychologically real, since speakers actively called upon them in making judgements about the origin of lexical items. To be sure, this evidence helps validate aspects of Kekchi grammar, namely its morpheme structure conditions. (For details, see Campbell 1974, 1976).

Perhaps more importantly, however, this external evidence shows that morpheme structure conditions are real, and cannot be accounted for merely by syllable structure rules as proposed by "Natural Generative Phonologists" (Hooper 1975). Thus external evidence provides the means for testing theoretical claims. External evidence has been shown to have important implications for several issues in linguistic theory, e.g. the controversy over

extrinsic ordering of rules, abstractness, morpheme structure conditions, etc. (See Campbell 1974, 1976, 1977; Kiparsky 1974.)

To conclude, psychological reality can be investigated empirically but it takes more than ransacking a body of data for the internal patterns and processes a linguist might find. It requires that evidence outside these internal patterns be sought which shows speakers using the rules of their language productively. As more and more cases of external evidence are considered, important issues in phonological theory may be resolved, and the answers to important questions found, questions such as: how different from the surface may underlying forms be and still be learned by speakers?; how many forms must illustrate a rule before speakers learn the rule rather than the variant forms piecemeal?; how do exceptions, non-productivity, non-phonetic conditioning factors, "opacity", and the like affect the learnability of rules?, etc. To answer these and related questions, we need sufficient external evidence, and until we answer them, phonological theory will be found wanting.

References

- Campbell, L. (1974): "Theoretical implications of Kekchi phonology", *IJAL* 40, 59-63.
- _____. (1976): "Linguistic acculturation: a cognitive view", In: *Studies in Mayan linguistics 1*, M. McClaran (ed.), American Indian Culture Center, UCLA.
- _____. (1977): "Generative phonology vs. Finnish phonology: retrospect and prospect", In: *Papers from the Transatlantic Finnish Conference*, R. Harms and F. Karttunen (eds.), 21-58, Texas Linguistics Forum, 5. Austin.
- Fromkin, V. (1971): "The non-anomalous nature of anomalous utterances", *Lg.* 47, 27-52.
- _____. (1973): *Speech errors as linguistic evidence*, (Janua Linguarum, Series Maior, 77), The Hague: Mouton.
- Hooper, J. (1975): "The archi-segment in natural generative phonology", *Lg.* 51, 536-60.
- Kiparsky, P. (1968): "Metrics and morphophonemics in the Kalevala", In: *Studies presented to Professor Roman Jakobson by his students*, C. Gribble (ed.), 137-48, Cambridge, Mass.: Slavica.
- _____. (1972): "Metrics and morphophonemics in the Rigveda", In: *Contributions to generative phonology*, M. Brame (ed.), 171-200, Austin: University of Texas Press.
- _____. (1974): "On the evaluation measure", In: *Papers from the parassession on natural phonology*, A. Bruck, R. Fox and M. LaGaly (eds.), 328-37, Chicago Linguistic Society.

- Ohala, M. (1974): "The abstractness controversy: experimental input from Hindi", Lg. 50, 225-35.
- Sherzer, J. (1970): "Talking backwards in Cuna: the sociological reality of phonological descriptions", Southwest Journal of Anthropology 26, 343-53.
- Zeps, V. (1963): "The meter of the so-called trochaic Latvian folksongs", International Journal of Slavic Linguistics and Poetics, 7, 123-8.
- _____. (1973): "Latvian folk meter and styles", In: A Festschrift for Morris Halle, S. Anderson and P. Kiparsky (eds.), 207-11. New York: Holt.

THE PSYCHOLOGICAL REALITY OF WORD FORMATION AND LEXICAL STRESS RULES

Anne Cutler, Experimental Psychology, University of Sussex, England

Introduction

'Psychological reality' has both a strong and a weak sense. In the strong sense, the claim that a particular level of linguistic analysis X, or postulated process Y, is psychologically real implies that the ultimately correct psychological model of human language processing will include stages corresponding to X or mental operations corresponding to Y. The weak sense of the term implies only that language users can draw on knowledge of their language which is accurately captured by the linguistic generalisation in question. For certain linguistic constructs, this weak sense embodies no more than a claim to descriptive adequacy; for example, the intuitions which the weak reading of 'psychological reality of the phoneme' predicts speakers will show are the same distributional data which led to the postulation of such a construct in the first place. This is not true of transformational rules - even to claim the weak sense of psychological reality for these is to claim that speakers can draw on knowledge at some level of the structures preceding and following application of the rule.

Lexical stress rules and word formation rules are transformational in nature. Within the grammar, the former are generally assumed to comprise part of the phonology, whereas the latter are claimed by some (Aronoff 1976) to constitute a separate stage preceding application of all phonological rules.

I wish to argue that the available evidence suggests psychological reality in the weak sense for both types of rule, as currently formulated in linguistic theory, but psychological reality in the strong sense for neither. (Note that this argument cannot be generalised to other phonological descriptions; see Fromkin (1973) for an argument in favor of strong psychological reality of abstract phonological representations).

Lexical Stress Rules

I have previously argued (Cutler 1977) that speech error evidence does not suggest the application of lexical stress rules in the production process, i.e. that lexical stress errors do not exemplify the misapplication of stress rules. What might we expect from an error in stress rule application? Fay's (1977a) argument for the strong psychological reality of syntactic transformations

is based on errors which Fay claims show that a particular rule (a) has failed to apply (what he said? for what did he say? is analysed as failure to apply Subject-Auxiliary Inversion), or (b) has applied only partially (Do I have to put on my seat belt on? is explained as application of the movement but not the deletion involved in Particle Movement). Since the function of lexical stress rules is to assign greater relative prominence to one syllable in a word than to others, one might expect that either failure to apply the appropriate rule or only partial application would result in less than the expected difference in degree of prominence between the syllables of a word. That is, if no stress rule applied at all one might expect all vowels in the word to be (equally) prominent, or, possibly, (equally) non-prominent; if, say, the Stress Adjustment Rule failed to apply one might expect a syllable to bear tertiary stress when it should be unstressed, etc.¹ But in fact lexical stress errors result always in primary word stress falling on the wrong syllable, not in lack of differentiation between syllable stress levels. Failure to apply the Alternating Stress Rule (Chomsky and Halle 1968: 78) would indeed result in stress falling on a wrong syllable, e.g. the third syllable of nightingale; but my corpus of lexical stress errors contains not a single such example.

A more complicated hypothesis could be proposed in which, for example, final consonants were misidentified, or the syllables in the word counted wrongly, so that stress ended up on the wrong syllable. But this hypothesis, like the hypothesis that a rule has not applied, in no way predicts the most striking characteristic displayed by lexical stress errors. This is that the syllable which wrongly bears stress is always a syllable which bears stress in another word with the same item. Typical errors are: economist (cf. economic); photographing (cf. photography); conflict_N (cf. conflict_V); disadvantageous (cf. disadvantage).

An explanation of these errors which does account for this curious regularity is the following: derivationally related words are in some sense stored together in the mental lexicon, with each word's individual specification including inter alia an indication of stress pattern (stressed syllable); a stress error occurs when

-
1. Such errors do occur, but only when another word derived from the same base has the intruding stress pattern; e.g. [djúpIkèt] for [djúpIkət].

the stress syllable marking selected is not the one belonging to the target word, but that belonging to one of the other words in the group. (This explanation also accounts for the second, corollary, regularity exhibited by stress errors: they occur only in derived words and only in members of the Latinate section of the English vocabulary. The Germanic section of English is much less rich in morphologically related pairs of words with different syllables stressed, hence it provides less often the necessary conditions for occurrence of a lexical stress error).

It is clear that this explanation, by assuming stress pattern to be marked in the lexicon, implies that lexical stress rules do not apply in the course of language production.

However, there would seem to be no doubt that English speakers can draw on knowledge about the principles governing stress assignment in their language. Many experimental studies (e.g. Ladefoged and Fromkin 1968; Trammell 1978) have found that subjects' pronunciations of non-words or unfamiliar words conform fairly well to the predictions of the lexical stress rules; although Nessly (1977) used similar data collection methods to adduce evidence in favor of his own version of the rules rather than Chomsky and Halle's. Since language users normally find little difficulty in the task of assigning lexical stress in unfamiliar words, names and nonsense words, some representation of the principles underlying English stress assignment must be available to them, i.e. something more abstract than the mere aggregate of all the stress markings stored for all the individual words in their lexicon.

Word Formation Rules

Aronoff (1976:22, 46) and Halle (1973:16) specifically exclude any claim to psychological reality of word formation rules in the strong sense. Nevertheless there is evidence from speech errors which could be interpreted as favoring such a claim. Admittedly, one hardly ever finds errors in which a word formation rule seems to have failed to apply, i.e. substitution for the target word of the word or morpheme (depending on one's formulation of the rules) which formed the base of the target - say, familiar for familiarity; for one thing, preservation of target form class is one of the strongest characteristics of word substitution errors of any kind (Fromkin 1973; Fay and Cutler 1977). But errors do occur in which the wrong ending, albeit one appropriate to the form class, is produced: derival for derivation (Fromkin 1977), self-indulgement

for self-indulgence. A possible interpretation of these errors is that the wrong word formation rule has been applied.²

It will be obvious, however, that the model suggested in the previous section excludes the application of word formation rules in production as firmly as it excludes the application of lexical stress rules; if word formation rules operate, stress could not be marked in the lexicon as it would be dependent on the operation of the word formation rules. Can this model assign an interpretation to the suffix errors mentioned above? One obvious remark to be made about these errors is their similarity to prefix errors as discussed by Fay (1977b). Prefix errors result in one prefixed word substituting for another (e.g. intention for attention) or a non-word being formed by the addition of an inappropriate prefix (concustomed for accustomed). Similarly suffix errors can result in non-words (e.g. likeliness for likelihood) or in words (necessitous for necessary; these latter errors, word substitutions in which target and error differ only in the suffix, are of course difficult to distinguish from semantic errors and malapropisms). Fay suggested that prefixed words with the same stem might be stored together in the lexicon, and a prefix error result when not the target prefix but a neighbouring prefix was selected by mistake. It is clear that a similar proposal could account for suffix errors producing real words. Thus the lexical entry for a word family would be headed by the stem; the detailed entry for each member of the family would specify affixes, if any, number of syllables (see Engdahl (1978)) and an indication of which syllable should bear lexical stress. To account, however, for both prefix and suffix errors which produce non-words, the model needs to be extended, perhaps to allow the production device to select an appropriate affix from its affix inventory in cases in which the target affix became in some way momentarily unavailable. (It is noteworthy that even when an affix error includes a stress error, stress in the error occurs on a syllable which bears stress in some member of the word family.) To propose factors which might precipitate affix unavailability, i.e. which might render the affix temporarily difficult for the production device to interpret, is, however, to enter the realm of pure

2. These errors show no general tendency for affixes with + or # boundaries to prevail, or for more productive affixes to replace less productive.

speculation. It is to be hoped that more light will soon be shed on this issue; for the time being we must acknowledge that the evidence does not strongly support any particular model.

There is no doubt at all, however, that the facts of word formation have a claim to psychological reality in what we have identified as the weak sense. All the speech error evidence which has been discussed above and which has been interpreted as support for a model of the mental lexicon in which related words are stored together also provides clear support for the psychological reality of morphological structure. A considerable body of psycholinguistic evidence also supports this conclusion (e.g. Taft and Forster 1975). Whether or not rules of word formation of the particular type proposed by Aronoff are available to English speakers to generate new and nonce words is however uncertain. Aronoff and Schvaneveldt (1978) report that subjects in a lexical decision study are more likely to produce false positive responses to non-words formed with the productive suffix -ness than with the less productive suffix -ity, a result predicted by Aronoff's model.

However the results of an informal study of my own were less clearcut. In this study subjects were asked to choose between two candidates for words to fill what amounted to a gap in the language (e.g. to choose between excusal and excusement for 'act of excusing'); each pair of neologisms comprised one word formed with a # boundary (-ness, -ment, -ise, -ish, -y) and another formed with a + boundary suffix (the latter, which often result in stress falling on the suffix rather than on the stem, are considered to be less productive than the # boundary suffixes). Many of the words used were listed in the OED, but none in the Concise Oxford Dictionary, and in fact none of the 12 subjects, graduate students and faculty in psychology and language, claimed to recognise any word.

Since I used only 24 pairs and made no attempt to cover all possible combinations the results can hardly be considered conclusive. Nevertheless some interesting tendencies came to light. In general, subjects showed approximately equal preference for the more and the less productive endings. All subjects preferred excusal to excusement and despisal to despisement, although the OED lists all 4 forms; similarly, subjects preferred amassal and adressal although the OED lists only amassment and addressment. -ness was preferred to -ity for sinister (OED lists both sinisterity and sinisterness for 'quality of being sinister') and incestuous (OED: -ness only),

but accidentality was preferred to accidentalness (OED has both). For verb formations subjects seemed not to be able to make confident choices, and no clear trends emerged; an indication of the confusion can perhaps be seen in the fact that whereas more subjects preferred rapidify to rapidise for 'make rapid', vapidise was chosen more often than vapidify for 'make vapid'. Adjectives revealed yet another pattern of results in that subjects formed two clear groups, those who consistently preferred the less productive + affixes and chose, e.g., spectatorial, plumageous, and dowagerial, and those who consistently chose the more productive # affixes, i.e. spectatorish, plumagy, dowagerish.

The most that can be extracted from these findings is the conclusion that English speakers do not exhibit a great degree of unanimity in their choice of nonce formations. However some light is shed on the psychological reality of word formation processes by a comment made by several subjects independently, namely that although words formed with the + affixes (-al, -ity, -ify, -ial, -ous) were aesthetically more pleasing and would be preferred as permanent additions to the vocabulary, a # affix would generally be more useful to achieve understanding in everyday conversation. Thus although villagerial might in general be preferable to villagerish as an English word, the latter would be more likely to get the message across to an audience not expecting an unfamiliar word. Words with # affixes, which leave stress on the stem, are in other words recognised by speakers to be morphologically more transparent.

Conclusion

Morphological structure is psychologically real in that English speakers are aware of the relations between words and can form new words from old. The principles underlying lexical stress assignment are psychologically real in the sense that speakers know the stress pattern of regularly formed new words. The extent to which such knowledge proceeds from competence in the language or awaits conscious insight into morphological relationships is however unclear. It has frequently been suggested to me that morphological influences apparent in my stress error corpus results from error collection within a highly literate and linguistically sophisticated population. If so, then a speaker of English who knows, for example, the words economic and economist but is unaware of any relation between them should presumably not produce a stress error involving either of them. There is certainly no reason why

the structure of the mental lexicon should not be altered as a result of new knowledge about word structure being incorporated in the form of newly set up groupings or connections. But it is also possible that we know more than we are aware of. Recall Fay's discussion of prefixed words; how many of us are consciously aware, for example, that the stem spect in respect appears also in expect? It is at least possible that our mental lexicon could contain such knowledge even if we were not capable of making conscious use of it.

References

- Aronoff, M. (1976): Word Formation in Generative Grammar, Cambridge, Massachusetts: MIT Press.
- Aronoff, M. and R. Schvaneveldt (1978): "Testing morphological productivity". Unpub. MS.
- Chomsky, N. and M. Halle (1968): The Sound Pattern of English, New York: Harper and Row.
- Cutler, A. (1977): "Errors of stress and intonation", paper presented to 12th International Congress of Linguists, Vienna.
- Engdahl, E. (1978) "Word stress as an organizing principle for the lexicon". Papers from the Parasession on the Lexicon, Chicago Linguistic Society.
- Fay, D. (1977a): "Transformational errors", paper presented to 12th International Congress of Linguists, Vienna.
- Fay, D. (1977b): "Prefix errors", paper presented to 4th International Linguistics Meeting, Salzburg.
- Fay, D. and A. Cutler (1977): "Malapropisms and the structure of the mental lexicon", Linguistic Inquiry 8, 505-520.
- Fromkin, V. (1973): Speech Errors as Linguistic Evidence, The Hague: Mouton.
- Fromkin, V. (1977): "Putting the emphasis on the wrong syllable", in Studies in Stress and Accent, L.M. Hyman (ed.), Los Angeles: USC Press.
- Halle, M. (1973): "Prolegomena to a theory of word formation", Linguistic Inquiry 4, 3-16.
- Ladefoged, P. and V. Fromkin (1968): "Experiments on competence and performance", IEEE Transactions on Audio and Electroacoustics, 16, 130-136.
- Nessley, L. (1977): "On the value of phonological experiments in the study of English stress", in Studies in Stress and Accent, L.M. Hyman (ed.), Los Angeles: USC Press.
- Taft, M. and K. Forster (1975): "Lexical storage and retrieval of prefixed words", JVLVB 14 638-647.
- Trammell, R.L. (1978): "The psychological reality of underlying forms and rules for stress", J. Psycholing. Res. 7, 79-94.

PROBLEMS IN ESTABLISHING THE PSYCHOLOGICAL REALITY OF
LINGUISTIC CONSTRUCTS

Bruce L. Derwing, Department of Linguistics, University of Alberta,
Edmonton, Alberta, Canada T6G 2H1

The "psychological reality" issue in linguistics - and in phonological theory in particular - has many more facets than seem to be generally recognized. The first problem is the recognition that a problem in fact exists. Under the influence of ideas which developed originally in comparative philology, the prevailing linguistic philosophy has long been one of autonomy: a language has been viewed as a kind of isolated "natural object" which could be investigated independently of the psychology of its speakers and hearers. In recent years, this misapprehension has led to a concept of linguistic "competence" (Chomsky 1965) which consists of nothing more than an arbitrary set of "coding principles" (Straight 1976) abstracted by linguists from linguistic data and treated as something quite distinct from the mechanisms of listening and speaking. Yet, in fact, a language is not an isolated "thing" at all, but is rather the product of various psychological and physiological processes which take place within human beings. Physically, the language product can be studied in the form of speech articulations, acoustic waves, or peripheral auditory events, but in none of these three observable, physical states can we find anything which smacks of linguistic "structure" (not even "phones," which already involve considerable processing by the human perceptual apparatus). Linguistic "structure," therefore, if this term refers to anything real at all, must refer to representations or interpretations imposed upon the speech signal by language users, normally as part and parcel of the communication event itself (Derwing 1973, 302-307). In short, psychological reality is not merely a convenient luxury which linguistic theory may or may not choose to be concerned with, but is rather the sine qua non for any linguistic construct which aspires to anything more than an epiphenomenal or artifactual status, and hence for any linguistic theory which can justifiably claim to go beyond the bounds of an arbitrary taxonomic system.

It is for this very reason, in fact, that all of modern "autonomous" linguistics suffers from an insoluble non-uniqueness

problem: any set of language forms can be correctly (i.e., accurately) described in many different ways, as even the simple example of the English plural inflection clearly shows (Derwing 1974). This implies that pure "linguistic" or "internal" evidence (i.e., evidence about "static" language forms, etc.) is quite inadequate to distinguish a wide range of theoretical alternatives. The only apparent solution to this problem (apart from the adoption of arbitrary principles for grammar "evaluation") is to redefine the nature of the discipline: to say that the goal of linguistics is not merely to describe utterance forms, but rather to describe the knowledge and abilities which speakers have to produce and comprehend them. Linguistic claims now become subject to the test of truth: whereas the forms will admit of numerous possible descriptions, there are many psychological claims about what the speaker knows or does which can be shown to be wrong or inadequate. So an expanded domain of "psycholinguistic" evidence can help to sort out alternatives which the traditional kinds of "linguistic" data could not.

To recognize the need to psychologize linguistics is one thing, however, and the actual practice is something else again. Chomsky himself declared linguistics to be a branch of cognitive psychology a full decade ago (1968), yet he and his followers still continue to embrace many of the same old "pernicious ideas" (McCawley 1976) which prevent this conception from becoming anything more than a slogan. In other words, while the so-called "Chomskyan revolution" may well have entailed a terminological re-orientation in the direction of the psychologization of linguistic jargon and associated claims, no corresponding methodological revolution accompanied these changes, with the result that the generative grammarians "continued to practice linguistics as it has always been standardly practiced" (Sanders 1977, 165). Such linguists may thus claim to seek or establish "psychological reality," yet they still persist in evaluating their theories on the basis of various "simplicity" considerations rather than on the basis of independent psychological evidence (as if the more general theory were, in fact, the most psychologically "valid"; contrast Fromkin 1976, 56, with Steinberg 1976, 385-386.)

But we are still merely scratching at the surface of the problem. It has become commonplace nowadays to find exhortations

in the linguistic literature to "expand the data base," often in much the same directions as outlined above, yet Greenbaum seems quite justified in expressing the doubt "whether linguists will abandon a particular linguistic formulation on the basis of psycholinguistic evidence" (1977, 127). Why should they? For, after all, since most linguistic theorizing was done within the non-psychological or "autonomous" linguistic tradition, it is seldom clear what particular psychological claims, if any, are to be associated with any particular linguistic analysis. Obviously, before we can ever hope to make use of new kinds of evidence to test or evaluate psychological claims, we must first know what the particular claims are that we are required to evaluate.

This is the crux of what I have called the interpretation problem for grammars (Derwing 1974). If grammars merely describe utterance forms, then evidence about such forms is the only kind relevant to the evaluation of grammars, and a selection from among competing grammars can only be made on the basis of criteria which are ultimately arbitrary. But if grammars relate in any way to psychological events or states, then we need to interpret grammars psychologically so as to make it clear what the new empirical implications of these grammars are. In other words, a formal grammar requires a psychological interpretation before it can become part of a psychological theory, and it is only the combination of the grammar plus the interpretation which can be put to an experimental test.

Now the problem of interpretation is not nearly so severe with respect to some of the older, more concrete linguistic notions as in the case of many of the more recent, abstract developments. In Derwing & Baker (1977), for example, a summary is provided of various straightforward interpretations, relevant tests, and new experimental data which help to answer the question of which, if any, of several obvious ways of describing the English plural inflection is psychologically real. Serious problems arise, however, when we come to analyses of the type discussed in Anderson (1974, 54-61), which involve the positing of single "underlying" lexical representations and "extrinsically ordered" phonological rules. Even ignoring the major problem of what psychological interpretation to place upon the general notion of the grammatical "generation" of forms (cf. Crystal 1974, 303), what psychological

sense can possibly be made, first of all, of a notion of rule "ordering" which has no relation to real time? To my knowledge, no one has ever even proposed a sensible real-world analogue of this idea, and without an interpretation, to repeat, it is impossible to tell what kind of experimental test is even relevant to evaluate its validity. Fortunately, in this instance, at least, the concept is one that linguistic theory seems to be able to get along very nicely without, merely by reformulating all rules in such a way that no arbitrary ordering relations are required among them (cf. Derwing 1973 & 1975). But we are still left with the problem of what psychological content we can associate with the linguist's notion of the "underlying" or "base" form in phonology. A few suggestions have at least been made in this case (e.g., Linell 1974; Ingram 1976; Birnbaum 1975), but none of them have yet seemed compelling enough for anyone to risk taking one out onto an experimental limb. There is, in any event, another, less direct route which can be taken in connection with this particular evaluation problem. The keystone argument is that there is no basis for positing a single "underlying" lexical representation for any set of supposed "morpheme alternants" unless the alternants in question can indeed be shown to represent the "same morpheme" for speakers. Thus a test which assesses a speaker's ability to "recognize morphemes" can indirectly provide evidence relevant to the question of the extent to which psychological theories might plausibly be constructed which incorporate the linguistic notion of the "underlying" form. For example, on the basis of "morpheme recognition" data collected by means of tests described in Derwing (1976), there is reason to believe that typical speakers judge a word-pair to contain a common morpheme only if the two words involved share a certain "critical" degree of both semantic and phonetic similarity, as independently assessed (Derwing & Baker in press). On this evidence, therefore, any linguistic analysis which posits a common lexical representation for words such as fable and fabulous, which lie outside of this "critical" area, is not even psychologically feasible for more than a very small minority of speakers.

While the recognition and solution of the interpretation problem represent, I think, the main barrier to the establishment of the psychological reality of linguistic constructs, there are still

quite a number of smaller obstacles which also have to be faced and overcome. For one thing, we must learn to resist the temptation to be "bathtub experimentalists" (i.e., prone to the cry of "Eureka!"). For even an investigator who fully recognizes the need both to interpret and to test linguistic theories on psychological territory may well (for lack of laboratory experience, for example) fail to anticipate many of the difficulties which can arise out of the very activity of devising, carrying out, and finally evaluating experiments. The most insidious of these difficulties, no doubt, is the one associated with the experimental artifact. For just as (autonomous) linguistic theorizing has yielded many concepts which have no real-life analogues in the knowledge or skills of real language users, so a particular experimental technique can also yield data which are more representative of the technique (or of his subjects' ingenuity) than of the subjects' control of the phenomenon of interest. A particular experiment does not always test in practice what the experimenter thinks it is testing in theory. I have encountered this problem at least twice in my own research (cf. Derwing 1976, 43-50) and Fromkin (1976) properly takes a few experimenters to task for perhaps jumping too fast to conclusions because of it. But in the last analysis there is only one sure way to dispel doubts about the "experimental artifact" and that is via the very painstaking route of cross-methodological verification: each evaluation problem must be approached by means of a variety of alternative experimental routes, in order to insure that the results obtained are independent of any particular experimental procedure.

There are, of course, other methodological problems to be mentioned, as well. There is always, for example, the possibility of the "just plain goof" whenever experimental data are collected, interpreted and evaluated, a danger that springs from causes as trivial as the mispunching of data cards to others as abstruse as failure to attend to assumptions which underlie a particular statistical model. Yet the most common type of error to sneak through a data analysis unattended, perhaps, is the one that results from a failure to take due cognizance of uncontrolled confounding variables (cf. Derwing & Baker 1977, 100), with the result that one's interpretation may be based on an apparent cause rather than the real one. But, again, there is no sure or simple formula to

guarantee safe passage through such treacherous and unpredictable waters as these; one can only take the utmost care possible in his own work, then hope that his readers and critics will pick out whatever errors and oversights may remain.

Finally, there is also the problem of the extraneous or "nuisance" variable, so called, no doubt, because it is often so very hard to eliminate from the experimental situation, even when the investigator may know full well that it is there. In my own "morpheme recognition" research, for example, the interpretation of the data is continually muddled by the factor of possible orthographic interference. How much are "linguistic intuitions" conditioned by the academic task of learning how to read, thereby complicating our efforts to understand the "natural" course of language acquisition through mere exposure to spoken language forms under normal circumstances of use? (A very similar question is the one concerning the very validity of the "linguistic intuitions" of subjects who have already been exposed to any significant degree of formal linguistic training; cf. Derwing in press.) Answers to such questions can only be partially and very tentatively answered so long as one is forced to deal with literate (or "non-naive") experimental subjects. I am very happy to see, therefore, that some aspects of my work are soon to be replicated and extended to the study of Lapp morphology by R. Endresen of the University of Oslo, for included in his population samples will be many speakers who are not only linguistically untrained, but also illiterate in their own language, thereby making it possible to investigate systematically at least some effects of the orthographic variable. Unfortunately, not all "nuisance" factors can be so conveniently dealt with, and these others will continue to constitute one of the more troubling aspects of trying to advance our knowledge by means of controlled experimental research. But since this is the way of science and the only secure route we know of for establishing knowledge about the world and its inhabitants, we have little real choice but to face them all head on.

References

- Anderson, S.R. (1974): The organization of phonology, New York: Academic Press.
- Birnbaum, H. (1975): "Linguistic structure, symbolization, and phonological processes", in Phonologica 1972, W.U. Dressler et al. (eds.), 131-143, Munchen & Salzburg: Wilhelm Fink.

- Chomsky, N. (1965): Aspects of the theory of syntax, Cambridge, Massachusetts: MIT Press.
- Chomsky, N. (1968): Language and mind, New York: Harcourt Brace & World.
- Crystal, D. (1974): Review of R. Brown, A first language. Journal of Child Language 1, 289-334.
- Derwing, B.L. (1973): Transformational grammar as a theory of language acquisition, London: Cambridge University Press.
- Derwing, B.L. (1974): "English pluralization: a testing ground for evaluation", to appear in Experimental linguistics, G.D. Prideaux et al. (eds.), Ghent: E. Story-Scientia.
- Derwing, B.L. (1975): "Linguistic rules and language acquisition", Cahiers Linguistiques d'Ottawa, No. 4, 13-41.
- Derwing, B.L. (1976): "Morpheme recognition and the learning of rules for derivational morphology", Canadian Journal of Linguistics 21, 38-66.
- Derwing, B.L. (in press): "Against autonomous linguistics", in Evidence and argumentation in linguistics, T. Perry (ed.), Berlin & New York: de Gruyter.
- Derwing, B.L. & W.J. Baker (1977): "The psychological basis for morphological rules", in Language learning and thought, J. Macnamara (ed.), 85-110, New York: Academic Press.
- Derwing, B.L. & W.J. Baker (in press): "Recent research on the acquisition of English morphology", in Studies in language acquisition, P.J. Fletcher et al. (eds.), London: Cambridge University Press.
- Fromkin, V.A. (1976): "When does a test test a hypothesis?", in Testing linguistic hypotheses, D. Cohen et al. (eds.), 43-64, New York: Wiley.
- Greenbaum, S. (1977): "The linguist as experimenter", in Current themes in linguistics, F.R. Eckman (ed.), 125-144, New York: Wiley.
- Ingram, D. (1976): "Phonological analysis of a child", Glossa 10, 3-27.
- Linell, P. (1974): Problems of psychological reality in generative phonology, Uppsala University, Department of Linguistics.
- McCawley, J.D. (1976): "Some ideas not to live by", Die neueren Sprachen 75, 151-165.
- Sanders, G.A. (1977): "Some preliminary remarks on simplicity and evaluation procedures in linguistics", in L.G. Hutchinson (ed.), Minnesota Working Papers in Linguistics and Philosophy of Language, No. 4, 155-167.
- Steinberg, D.D. (1976): "Competence, performance and the psychological invalidity of Chomsky's grammar", Synthese 32, 373-386.
- Straight, H.S. (1976): "Comprehension versus production in linguistic theory", Foundations of Language 14, 525-540.

ARGUMENTS AND NON-ARGUMENTS FOR NATURALNESS IN PHONOLOGY:
ON THE USE OF EXTERNAL EVIDENCE

Wolfgang U. Dressler, Institut für Sprachwissenschaft,
University of Vienna, Austria

§1.0 The concept of naturalness has become a major concern for many phonologists. In my view, the concept of naturalness should be best regarded as a basic principle of a phonological theory and should be tested by the judicious use of external (or substantive) evidence.

As to the relationship of naturalness to psychological reality, my point is that a natural phonological analysis of a phenomenon claims psychological reality for its concepts and constructs. However, not all psychologically real constructs in a phonological analysis need to be phonologically natural. E.g., a phonological process (henceforth PR) posited by the linguist may refer to constructs of natural morphology (cp. Mayerthaler to appear), especially in case of so-called morphological rules (cp. Dressler, 1977a).

§1.1 In the theory of Natural Phonology (henceforth NatPhon), as proposed by Stampe since 1968 (see now Donegan and Stampe to appear) and 'Polycentristic Phonology' (Dressler 1977a), naturalness occupies a central place. Phonological systems are phonetically (and I add, psychologically and, to a lesser degree: sociologically, historically) motivated. The basic constructs of Natural Processes in the sense of "mental substitutions which systematically but subconsciously adapt our phonological intentions to our phonetic capacities" (Donegan and Stampe to appear, §1, including its perceptive converse) are substantive universals.

§1.2 Similar to adherents of NatPhon, S. Schane and M. Chen (see Sommerstein 1977, 230, 233) have claimed that particular languages select PRs from a fixed universal set of natural processes and may impose constraints on their applicability. In the best of cases a PR forms a subset of a universal process (as characterized by the theory) and any restrictions vis-a-vis the general form of the respective universal process can be derived from the hierarchies of the universal process and from a fairly small number of principles of restrictions.

But what if a PR is not such a regular subset of a universal process? In this case NatPhon (or at least Polycentristic Phonology) cannot appeal to frequency or intuitive plausibility, but

must explain why the given PR is not a regular subset of a universal process. Several avenues are open: 1) Modification of the universal process. 2) The deviation is due to language acquisition; in this case well-motivated linguistic and psychological concepts must explain the deviancy. 3) The deviation is due to historical circumstances (including sociological factors). 4) The PR is not totally (phonologically) natural, a possibility avoided in NatPhon (but cp. Dressler 1977a; Sommerstein 1977, 235f). Since such PRs (diachronically) must go back to totally natural PRs, explanation 4) includes explanations 3) and 2).

§1.3 Thus, it becomes clear that external evidence, at least from language acquisition, diachrony, and sociolinguistics is not external for NatPhon, but forms an integral part of the area it has to cover. Moreover, there is no theoretical or methodological principle which should exclude other dimensions of external evidence from investigation:

§1.3.1 Take sociophonology: The restriction to the investigation of only one level of formal, maximally differentiated speech as practised in most of generative phonology and almost all of structural phonology is an undue limitation of interest and of access to natural speech, whose variation is apt to give important insights even to formal principles of rule application (cp. Dressler 1975). However, any detailed and theoretically sound work on casual vs. formal speech presupposes the inclusion of both, psychological/psycholinguistic theory (cp. Vanecek and Dressler 1977) and sociological/sociolinguistic theory (cp. Wodak and Dressler 1978).

§1.3.2 Or: The differential (and always non-random!) breakdown of phonology in aphasia gives important insights into the structure of phonology. However, studies so far have not completed the desirable integration of all disciplines relevant to aphasia, e.g. the brilliant thesis of Keller (1975) neglects all recent phonological theories, whereas the present writer's studies (Dressler 1977b; 1978) have not yet integrated neuropsychology. For other types of external evidence, see Linell (1974), Fischer-Jørgensen (1975, 224ff), Zwicky (1975), Skousen (1975).

§2. Non-arguments for naturalness

In the literature we find certain non-arguments/fallacies:

§2.1 "Facts about the real working of the brain are most important".

Anttila (1977, 221) believes to have found direct evidence, as opposed to indirect neurolinguistic evidence, against generative grammar, when he cites the biologist W. Wieser about the brain not working exactly, often blundering and correcting itself, not proceeding logically, but according to similarities, being extremely redundant, etc. However, Wieser has informed me that these phenomena at the micro-level do not preclude precise rules at the macro-level (which is the level of interest for linguistics), just as Heisenberg's indeterminacy relation does not vitiate the precise working of laws of classical physics in macrophysics. Here we might speak of a micro-anatomic fallacy.

§2.2 There is a similar fallacy which one might call the macro-anatomic fallacy or mistaken equation of phonology and phonetics, which is an exaggeration of the Physical/Phonetic Basis Condition (Botha 1978 II, 16ff) of phonology. This line of argument neglects the interaction between phonological and morphological or phonological and lexical naturalness (cp. Dressler 1977a) and of what Hyman (1977) has called phonologization (which in my view starts with allophonic PRs producing extrinsic instead of intrinsic allophones).

§2.3 Still more common is the false equation of naturalness with concreteness, since as a result of refusing the abstractness involved in standard generative phonology, many phonologists have regarded concreteness as a virtue in itself. However, phonological concreteness has often been achieved at the expense of morphology for which very few 'concrete phonologists' (e.g. Skousen 1974) have cared to provide a theoretical framework. More important still, concreteness has been defined (if at all) as restrictions on the relationship between underlying phonological and surface phonetic representations. In my opinion it is possible to define the naturalness of processes and of representations (be it as structural symmetries as found by phonemicists or natural asymmetries as derived from processes, see Stampe (1973)), but not the naturalness of relationships between representations. Notice both the failure of Kenstowicz and Kisseberth (1977) to find universal formal constraints on the distance between phonological and phonetic representations (cp. Gussman 1978, 154, 167f; Sommerstein 1977, 237 n. 47), and the undesirable results of the much more rigorous restrictions of Natural Generative Phonology (see Hooper (1976) and

its critique by Gussman (1978, chapter 1) and Donegan and Stampe to appear, §3.1., §4).

As an example I simply want to refer to the abstract analysis of German [ŋ] as underlying /ng/ (discussed in detail in Dressler to appear; cp. Dressler (1977a, 51)). For the much debated PR $g + \emptyset/\eta$ — (except before non-centralized vowel), I have found external evidence, e.g. in loan-word integration and sociophonological variation (e.g. [^lʌŋgɛla] vs. [^lʌŋɛla] 'Angela'), in child language (Mandarine 'tangerine' + [mʌŋgʌ'ri:nə] vs. [mʌŋɛ'ri:nə]) and aphasia (see Stark 1974). Thus, multiple external evidence has been found in support of the psychological reality of this PR, although this analysis implies a very abstract underlying representation (cp. Kenstowicz and Kisseberth (1977, 7f, 53), Gussman (1978, 168), see below §3.4).

§2.4 Often natural is falsely equated with productive. This equation (Fischer-Jørgensen (1975, 228f); Skousen (1975); Linell (1976), etc.) might hold most of the time, but not always (Dressler (1977a, 1977c)).

§2.5 Still weaker and never explicitly justified is the equation of natural and (e.g. typologically) frequent. Frequency might be a first indicator for the phonologist looking for universals, but what counts is explanation in the sense of causal argumentation.

§3. Counterarguments against external evidence

§3.1 "External evidence is unnecessary, internal evidence suffices". This 'Nonnecessity Thesis' has been proved by Botha (1978 II) to be incompatible with empirical mentalism. Formal, 'pure' linguistics cannot alone do the job of vouching for psychological reality. Due to the serious underdetermination of standard data (internal evidence), various sources of external evidence must be adduced (cp. §1.3 and Botha (1978 II, III §5.3)).

§3.2 "External evidence is too unclear". However, internal evidence based on intuitions as utilized in generative phonology is unclear itself in many respects as Ringen (1975) has shown. Moreover, it must be noted that evidence from diachrony and loan-word integration seems to be accepted by many who shun other external evidence.

Unfortunately, the use of both types of evidence has been grossly simplified by most generative phonologists; for loan-words see Fischer-Jørgensen (1975, 229), Kiparsky (1973, 112ff), Dressler

(1977a, 35ff). As to diachrony, both structuralists and generativists have limited themselves far too often to nomological explanations (e.g. symmetry, rule simplification), while neglecting the all-important genetic explanation, e.g. by confusing sound change with sound correspondences; thus context-free processes have been liberally adduced as evidence, although they are, I believe, always the final result of generalizing context-sensitive sound change.

§3.3 External evidence shows "what in fact counts as internal evidence" (Kenstowicz and Kisseberth 1977, 3). Does this mean that e.g. English loan-words in Japanese might be used to demonstrate the necessity of morpheme structure constraints or redundancy rules within phonological theory, but not for corroborating their specific forms in Japanese itself?

§3.4 "Internal evidence is more important than external evidence", a view held by many (called the Nonprivileged Status Thesis by Botha (1978 II, 12f)). However, quite apart from its theoretical shakiness (cp. §1), there are counterexamples: E.g. the abstract analysis of English [ŋ] as /ng/ rests on exceptional (and thus suspect) alternations like lo[ŋ], lo[ŋ]est, whereas the normal, productive superlatives are e.g. bori[ŋ]est, winni[ŋ]est; but external evidence for the abstract analysis is excellent (starting with Fromkin (1973, 223)). Even more extreme is the German situation, where in most varieties internal evidence is restricted to distributional evidence (Vennemann 1970), which generativists usually esteem much less than evidence from alternations, alternations in this case exist only in external evidence (see above §2.3).

§3.5 Botha (1970, 130ff) has deplored the 'qualitative type jump' from internal to external evidence and the lack of criteria of adequacy. Since then he has revised his standpoint and has demanded the construction of "bridge theories" mediating between linguistics and other disciplines relevant for the given type of external evidence (Botha 1978 III, 27ff). But 'hyphenated' disciplines, such as psycholinguistics, sociolinguistics, neurolinguistics have strived just for that since many years!

§3.6 "External evidence is often divergent and incoherent" (Gussman (1978, 167f) happily cites Dressler (1977d, 224), where higher standards in the use of external evidence are demanded). Here Botha (1978 III, 30, 27ff) correctly states "that the relative weight of a given kind of external evidence is a function of the

adequacy of a particular bridge theory". In other cases conflicting external evidence may force us to revise phonological theory (e.g. in the case of introducing Korhonen's concept of 'quasi-phonemes' in Dressler (1977a, 52ff).

§3.7 In connection with §3.6 I want to discuss a problem which seems to strike a heavy blow to the theory espoused here: I have linked naturalness firmly with the universality of natural processes. However, processes actually studied, show different hierarchies, both typologically and in external evidence (cp. Drachman 1977; Ferguson 1978), although hierarchies have been claimed to be an integral part of the universal processes constructed by NatPhon. Whereas Atomic Phonology has found a purely formal solution (criticized by Donegan and Stampe (1977)) to this problem, I want to come back to §1.2. The reactions of an individual to innate physiological and psychological restrictions are determined both by maturation and social environment. In this way I agree neither with (rather mystical) strong claims about innate universals (as in certain quarters of TG), nor with the arbitrariness of the outcome of societal constraints (as implied in marxist critiques of TG). Therefore (in Dressler 1977a) I have spoken only of universal tendencies (one type being universal processes) which necessarily conflict and must be compromised by the language learner: Thus, certain universal processes are suppressed either in the language as a whole or in certain domains of external evidence; or they are restricted in ways allowing different process hierarchies. Moreover, a typology of phonological processes must consider advances made in the theory of typology: e.g. ordering typologies may be multilinear (with branchings).

References

- Anttila, R. (1977): Rev. of Linell 1974, Lingua 42, 219-222.
- Botha, R. (1970): The justification of linguistic hypotheses, The Hague: Mouton.
- Botha, R. (1978): On the method of mentalism, Ms., University of Stellenbosch.
- Donegan, P.J. and D. Stampe (1977): "On the description of phonological hierarchies", in CLS Book of Squibs, S. Fox et al. (eds.), 35-38.
- Donegan, P.J. and D. Stampe (to appear): "The study of natural phonology", in Current Approaches to Phonological Theory D. Dinnsen (ed.), Bloomington: Indiana University Press.
- Drachman, G. (1977): "On the notion 'Phonological Hierarchy'", in Phonologica 1976, W. Dressler and O. Pfeiffer (eds.), Innsbrucker Studien zur Sprachwissenschaft, 85-102.
- Dressler, W. (1975): "Methodisches zu Allegroregeln", in Phonologica 1972, W. Dressler and F. Mareš (eds.), Munich: Fink, 219-234.
- Dressler, W. (1976): "Tendenzen in kontaminatorischen Fehlleistungen", Sprache 22, 1-10.
- Dressler, W. (1977a): Grundfragen der Morphonologie, Vienna: Österreichische Akademie der Wissenschaften.
- Dressler, W. (1977b): "Morphological disturbances in aphasia", Wiener linguistische Gazette, 14, 3-11.
- Dressler, W. (1977c): "Morphologization of phonological processes", in Linguistic Studies presented to J. Greenberg, A. Juillard (ed.), Saratoga: Anma Libri, 313-337.
- Dressler, W. (1977d): Rev. Skousen 1975, Lingua 42, 223-225.
- Dressler, W. (1978): "Phonologische Störungen bei der Aphasie", Badania lingwistyczne nad afazją, Warsaw: Ossolineum, 11-22.
- Dressler, W. (to appear): "External evidence for an abstract analysis of the German velar nasal", in Phonology in the 1970's D. Goyvaerts (ed.), Ghent: Story-Scientia.
- Ferguson, Ch.A. (1978): "Phonological processes", in Universals of Human Language, J. Greenberg (ed.), Stanford Univ. Press, 403-442.
- Fischer-Jørgensen, E. (1975): "Perspectives in Phonology", Annual Report of the Institute of Phonetics, University of Copenhagen 9, 215-235.
- Fromkin, V. (1973): (ed.) Speech errors as linguistic evidence, The Hague: Mouton.
- Gussman, E. (1978): Explorations in abstract phonology, Univ. of Lublin.
- Hooper, J. (1976): An introduction to natural generative phonology, New York: Academic Press.
- Hyman, L. (1977): "Phonologization", in Linguistic Studies presented to J. Greenberg II, A. Juillard (ed.), Saratoga, 407-418.
- Keller, E. (1975): Vowel errors in aphasia, Univ. of Toronto.
- Kenstowicz, M. and Ch. Kisseberth (1977): Topics in phonological theory, New York: Academic Press.
- Kiparsky, P. (1973): "Phonological representations", in Three dimensions of linguistic theory, O. Fujimura (ed.), Tokyo: TEC, 1-136.
- Linell, P. (1974): "Problems of psychological reality in generative phonology", Reports from Uppsala University, Department of Linguistics 4.
- Linell, P. (1976): "Morphonology as part of morphology", Phonologica 1976, 9-20.
- Mayerthaler, W. (to appear): Morphologische Natürlichkeit, Ms. TU Berlin.

- Ringen, J. (1975): "Linguistic facts", in Testing linguistic hypotheses, D. Cohen and J. Wirth (eds.), Washington: Hemisphere Public Corp., 1-42.
- Skousen, R. (1974): "An explanatory theory of morphology", Papers from the Parasession on Natural Phonology, CLS, 318-327.
- Skousen, R. (1975): Substantive evidence in phonology, The Hague: Mouton.
- Sommerstein, A. (1977): Modern phonology, London: Arnold.
- Stampe, D. (1973): "On chapter nine", in Issues in phonological theory, Kenstowicz and Kisseberth (eds.), The Hague: Mouton, 44-52.
- Stark, J. (1974): "Aphasiological evidence for the abstract analysis of the German velar nasal", Wiener linguistische Gazette 7, 21-37.
- Vanecek, E. and W. Dressler (1977): "Untersuchungen zur Sprechsorgfalt als Aufmerksamkeitsindikator", Studia Psychologica 19, 105-118. (A preliminary version in Wiener linguistische Gazette 9, 1975).
- Vennemann, Th. (1970): "The German velar nasal", Phonetica 22, 65-82.
- Wodak, R. and W. Dressler (1978): "Phonological variation in colloquial Viennese", Michigan Germanic Studies 4, 1, 30-66.
- Zwicky, A.M. (1975): "The strategy of generative phonology", Phonologica 1972, Dressler and Mareš (eds.), 151-168.

ABSTRACT PHONOLOGY AND PSYCHOLOGICAL REALITY

Edmund Gussmann, Institute of English, Maria Curie-Skłodowska University, Lublin, Poland

For a number of years now abstract phonological descriptions have come under attack from two different but often related quarters.¹ Firstly, it has been claimed that even within the broad framework of standard generative phonology less abstract solutions are often available; reinterpretations of the data have been achieved by suggesting that certain putative phonological contrasts are in fact morpho-lexical generalisations, i.e. morphologically and lexically rather than phonologically conditioned. Re-analysis or change of underlying representations has also been offered as a viable alternative to manipulating abstract segments and opaque rules. Finally, various modifications in the rule component have been shown to lead to less drastic departures from phonetic representations than those called for by (relatively) abstract positions. The drive towards concreteness seems to have culminated in the rise of so-called 'natural generative phonology' of Vennemann, Hooper and others although a whole range of more or less abstract views has continued to exist; in fact these radically concrete positions are coming under attack now even from those linguists who generally favour concreteness in phonology (cf. Goyvaerts 1978, 125-133). In any case, the type of criticism of abstract solutions that is normally based on evidence internal to the structure of the language cannot be meaningfully discussed without taking into account the grammar as a whole, and this is obviously precluded here. It can be safely assumed that less abstract solutions will be acceptable even to those linguists who favour abstractness in phonology if it can be shown that abstract interpretations are not necessary, i.e. that either the required generalisations can be made without recourse to the abstract machinery or else that the generalisations are in fact wrong and must be replaced by others. It is perhaps worth stressing that in order to evaluate such arguments and counter-arguments one must consider not just individual pairs of rules but rather the phonology as a whole; there has been far too much specula-

(1) The bibliography of the subject is vast and would require several pages. In this report I have restricted myself to just a few items which are directly relevant to the discussion.

tion based on scattered examples and even on inaccurate data.

The other line of attack on abstract positions has involved external evidence which has come to be known as substantive evidence. It has been claimed that the generalisations captured in abstract descriptions are not those that speakers of the language make, i.e. that the abstract generalisations are, in a nutshell, a figment of the linguist's imagination devoid of any psychological reality. This line stresses the need to go beyond the structural facts of the language in search of support for true generalisations. Substantive evidence for such psychologically real regularities has been sought in historical change, the treatment of borrowings, in language acquisition and language loss (aphasia), metrics, dialectal variation, speech errors, secret languages as well as in direct phonological experiments (see Fischer-Jørgensen 1975, 290ff and Zwicky 1975 for good surveys). These are important findings which certainly cannot be overlooked by anybody seriously concerned with psychologically real phonology. They must, however, be handled with extreme caution given the present understanding of the ways in which language is actually used since, as was judiciously observed by Dressler (1977, 224), "the more modalities of external evidence one uses, the more divergent and incoherent results one gets". Let me consider just a few cases.

Polish has a general and typologically very natural rule of devoicing obstruents word finally. In actual speech one often finds that the rule is suspended in certain cases, e.g. in regularly used foreign words and names whether completely assimilated into the language or not - gro[g] 'grog' rather than gro[k], ko[d] pocztowy 'postal code' (in spite of the fact that [d] precedes a voiceless plosive!), possibly because the unvoicing would produce here the humorous kot pocztowy 'postal cat'; in native words it is also suspended for a variety of reasons as in dó[b] '24 hrs., gen.pl.', where the unvoicing would produce a somewhat improper word. Surely no one would like to conclude from such examples that terminal unvoicing is not a psychologically real rule in Polish. Generally speaking, foreign words exhibit specific properties, and most schools of phonology have reflected this fact in one way or another (in addition it seems that one should also recognise varying degrees of foreignness). The fact that some foreign or occasional native words (including, possibly, nonsense words) do not appear

to have undergone a rule cannot be taken as direct evidence for the non-reality of the rule.

Historical evidence, one of the most important sources of substantive evidence, is notoriously difficult to handle in that the paucity or lack of reliable and unambiguous data is not the only factor hampering definite conclusions; any interpretation of change for purposes of verifying general theoretical claims involves assumptions about the mechanisms of change which themselves are not well understood and it also involves assumptions about e.g. the interface between the rules of morphology and those of phonology which is likewise largely unexplored. In view of these problems it is not surprising that examples can be found in the literature purporting to justify both abstract and concrete positions by use of such evidence. The metric evidence available from the works of Kiparsky, Anderson and others seems to support the level of remote representations although, given the variety of theoretical machinery accessible to current linguistic thinking, alternatives could presumably be found.

Slips of the tongue have figured prominently as the window to psychologically real grammars, and Fromkin's (1971) seminal paper has stimulated a lot of interest in this area. Some of her evidence has now become part of the stock-in-trade of those arguing for abstract regularities as, for example, the celebrated case for /ng/ as underlying the phonetic [ŋ]. It would be easy for somebody trying to defend abstract phonology to claim that if /ng/ underlies [ŋ] in a psychologically real sense, then speakers of English must have at their disposal means of arriving at the abstract solution given the data internal to the language. These means could then be generalised to cases where no external evidence can be adduced; this is the position adopted by Kenstowicz and Kisseberth (1977) who incidentally find that the case of the English velar nasal violates all of their constraints on the abstractness of underlying representations. Such evidence is intriguing, but supporters of concrete phonology could easily dispose of it by viewing the slips as resulting from the influence of spelling or something else. I would like to further emphasise, however, that important as such evidence may be, it is not obvious whether much use can be made of it until more is known about the interaction of linguistic knowledge and language use. In our particular case we need some sort

of theory of speech errors against which we could evaluate individual instances for their linguistic significance since one frequently observes not only slips of the tongue that can be shown to reveal something about the underlying reality of language but also instances of errors that appear to make "no sense" linguistically. It is also worth mentioning that different areas often provide contradictory evidence (cf. also Dressler's remark quoted above). The following might be a possible example: slips of the tongue adduced by Fromkin appear to suggest that affricates should be treated as single segments phonetically in English. On the other hand, optional low phonetic rules frequently simplify affricates to spirants in certain contexts so that French and orange end in [ʃ] and [ʒ]. This, of course, could be interpreted as a change in the feature /cont/ but since one also finds the deletion of alveolar plosives in such words as rents, sounds, it seems more plausible to treat both these changes as cases of deletion of the plosive between a nasal and a spirant. This would require, however, that affricates be clusters at some stage in the derivation.

The need for the study of the ways of utilising linguistic knowledge in speech is further confirmed by some surprising results obtained from direct phonological and grammatical tests. Earlier studies attempted to show that certain rules of the SPE phonology are not psychologically real as speakers fail to apply them to novel forms (nonsense words). Haber (1975) has shown that contrary to what might be expected speakers of English do very badly in tasks intended to test the productivity of the regular plural formation rule (the -(e)s ending), i.e. one that with good reason is generally assumed to be fully productive. It does not matter here whether the relevant mechanism is purely phonological, morphological or something else (the rule is transparent and could be formulated in surface terms). If tests fail to confirm the psychological reality of this simple rule, then most linguists would agree, I suppose, that there is something fundamentally wrong with the tests themselves; Kiparsky and Menn (1977, 64) ascribe it to "a "strangeness effect" which causes the subjects' performance to deteriorate relative to their normal speech" and are also (66-67) "skeptical about the ability of production tasks to show much of anything, at present, about the form of internalized linguistic knowledge, given the near-total obscurity surrounding the question of whether

and how this knowledge is used in speech".

As far as other areas of substantive evidence are concerned let me just mention two points: firstly evidence from an aphasiological study by Stark (1974) strongly suggests that the German velar nasal should be regarded as being derived from underlying /ng/, and this thus strengthens the case for an abstract interpretation of this problem vis-a-vis the stand taken by natural generative phonologists. Secondly, there is the case reported in Kiparsky and Menn (1977, 69-70) of an "invented language" which appears to exhibit two rules extrinsically ordered, which would indicate that the ordering of rules in itself cannot be difficult or impossible to learn as has been sometimes claimed. As Kiparsky and Menn point out, the charge that synchronic rule order mirrors diachronic developments cannot be made against speech invented by children.

The above discussion has not been meant to decry the importance of substantive evidence; conversely, in view of its potential significance I think it is necessary to stress that there is much in it which is arguable and which is itself in need of explanation and so can hardly be taken as definitive evidence for other theoretical concepts.

One final point that I would like to make is that the theoretical apparatus of abstract phonology is required to account for uncontroversially related, low phonetic details of pronunciation (see also Kiparsky 1975). Modifications, permutations, deletions and insertions of segments are well-known not only from abstract derivations but are also exceedingly common in accounts of rapid speech phenomena; thus, there is nothing basically new about abstract derivations that could not be found closer to the surface. Examples of the various modifications are well-known, and I would like to present a couple of examples from Polish where allegro rules introduce segments and contrasts totally absent from lento speech.² The phonetic inventory of Polish vowels contains six basic elements [i, ɨ, ɛ, a, ɔ, u], thus being again fairly regular typologically. Allegro forms introduce on the one hand a contrast of length which

(2) The examples are taken from Biedrzycki (1978) who interprets such data in terms of autonomous phonology and sets up phonemic distinctions for allegro styles which do not appear in lento styles.

does not appear in slow speech, e.g.: da 'she gave' [da:] vs. da 'she will give' [da], stó 'table' [stu:] vs. stu 'of a hundred' [stu] corresponding to the lento forms [dawa - da] and [stuw - stu], respectively, and also several segments which are not known elsewhere, e.g.: in sp[ə:] czeństwo 'society' cz[ə:]m 'hi' - lento sp[ɔwɛ] czeństwo, cz[ɔwɛ]m; zapomni[ə:]m 'I forgot', chci[ə:]m 'I wanted' - lento zapomni[awɛ]m, chci[awɛ]m; cz[o:] 'one felt', ok[o:] 'one shod' - lento cz[uwɔ], ok[uwɔ]. The low level, optional rules which produce such forms are psychologically real and by producing new contrasts they seem to work like absolute neutralisation in reverse. If we were to postulate length contrast phonologically for Polish and then absolutely neutralise it, the abstractness sin would be committed; speakers of the language, however, seem to find nothing unusual about neutralising certain contrasts and introducing new ones when passing from lento to allegro styles. The force of these examples should not be overstated but they seem to show that there is nothing abnormal about rules merging and producing contrasts or about segments which appear at one level of representation but not at another.

The abstractness debate will no doubt continue both on language internal and external grounds. There remains much to do in both areas so that any final verdict at this stage would be premature.

References

- Biedrzycki, L. (1978): Fonologia angielskich i polskich rezonantów, Warszawa: Państwowe Wydawnictwo Naukowe.
- Dressler, W.U. (1977): rev. of Skousen, Substantive evidence in phonology, Lingua 42, 223-225.
- Fischer-Jørgensen, E. (1975): Trends in phonological theory, Copenhagen: Akademisk Forlag.
- Fromkin, V.A. (1971): "The non-anomalous nature of anomalous utterances", Lg. 47, 27-52.
- Goyvaerts, D.L. (1978): Aspects of the post-SPE phonology, Ghent: E. Story-Scientia.
- Haber, L.R. (1975): "The muzzy theory", in Papers from the 11th Regional Meeting CLS, R.E. Grossman et al. (eds.), 240-256, Chicago: Chicago Linguistic Society.
- Kenstowicz, M. and Ch. Kisseberth (1977): Topics in phonological theory, New York: Academic Press.
- Kiparsky, P. (1975): "What are phonological theories about?", in Testing linguistic hypotheses, D.Cohen and J.R. Wirth (eds.), 187-209, Washington, D.C.: Hemisphere Publishing Corporation.

Kiparsky, P. and L. Menn (1977): "On the acquisition of phonology", in Language, learning and thought, J.Macnamara (ed.), 47-78, New York: Academic Press.

Stark, J. (1974): "Aphasiological evidence for the abstract analysis of the German velar nasal", Wiener Ling. Gazette 7, 21-37.

Zwicky, A. (1975): "The strategy of generative phonology", Phonologica 72.

THE PROBLEM OF PSYCHOLOGICAL REALITY IN THE PHONOLOGY OF PAPAGO

Kenneth Hale; M.I.T., Cambridge, Massachusetts, U.S.A.

This paper amounts to a clarification, for my own benefit more than anything else, of a certain linguistic principle involved in the evaluation of grammars. The principle, termed recoverability, is due to Jonathan Kaye (1975), and my conclusion about its nature and role in choosing among competing analyses is strongly influenced by a recent paper of Morris Halle's (1978). In examining the recoverability principle, I draw a comparison between two cases of alternative analyses - namely, the familiar problem of the Polynesian passive morphology, and a superficially similar problem in the phonology of Papago. In the course of the discussion, I will attempt to correct an error of reasoning which I made in an earlier treatment of the Polynesian case (Hale, 1973).

I begin with the Polynesian passive, using Maori to exemplify the classic situation. Some passives appear simply to involve straightforward suffixation of a vowel /-a/ - e.g., /patua/ beside active /patu/, /kitea/ beside /kite/, and so on. The majority, however, show a consonant between the root and a vocalic termination /-ia/ - e.g., /awhitia/ beside /awhi/, /hopukia/ beside /hopu/, /werohia/ beside /wero/, /inumia/ beside /inu/, etc. The consonant in these passives cannot be predicted from the surface form of the uninflected, or active, verb. There are at least two ways to analyze the consonantal passives. The 'phonological' analysis assigns the consonant to the stem, and the passive is formed by suffixing /-ia/ thereto. In addition, there is a rule deleting word-final consonants, thereby accounting for the fact that uninflected verbs, like all words in Maori, end in vowels. Thus, underlying /inum/ appears as [inu] if uninflected, but the deletion does not apply before the passive suffix, hence [inumia]. The 'morphological' analysis, by contrast, assigns the consonant to the suffix, thereby proliferating suffixal alternants (/ -tia, -kia, -hia, -mia.../), and each verb is assigned to a 'conjugation' according to the passive allomorph it selects. The assignment of the final consonant to forms like /inum/ is historically correct, and the deletion of these consonants in word-final position is a fact of the linguistic tradition leading to Polynesian. Nevertheless, I have attempted to argue (Hale, 1973) that the ahistorical morphological analysis is correct synchronically. If so, then what linguistic principle

dictates it? What motivated the change from the (historical) phonological analysis to the (ahistorical) morphological analysis? I suggested that the motivating factor was the canonical disparity, present in the phonological analysis, between the underlying morpheme structure of lexical items (allowing final consonants) and the surface syllabic canon of Polynesian (forbidding final consonants). The change to the morphological analysis eliminated this disparity.

Kaye has suggested another explanation. He defines a principle of 'phonological recoverability': "Recoverability concerns the degree of ambiguity manifested by a given surface form. The fewer the number of potential sources for the form, the greater its recoverability" (Kaye, 1975, 244-45). Phonological recoverability is valued in grammar, while 'phonological ambiguity', its converse, is devalued. Notice that the change in Polynesian completely eliminates phonological ambiguity - under the morphological analysis, the underlying form of a verb root is entirely recoverable from its surface form. Kaye proposes that phonological recoverability is the deciding factor in the Polynesian case.

This is a very promising suggestion. However, there is somewhat more texture to the problem which should be brought out in order to characterize the linguistic nature of the recoverability principle. As Halle (1978) points out, the Polynesian change did not really eliminate ambiguity. Rather, it shifted the ambiguity entirely to the morphology. The relation between uninflected and inflected forms remains ambiguous, since the derived form (the passive) is not predictable from the surface form of the active. Let us refer to this relation as 'morphological ambiguity' and to the converse relation as 'morphological recoverability'.

While phonological recoverability is logically distinct from morphological recoverability, it is not at all clear that the two principles are linguistically distinct. At least, I know of no convincing case in which phonological recoverability can be said to function autonomously in the evaluation of grammars. If the Polynesian change had consisted solely in the restructuring (i.e., in the realignment of the historic stem-final consonants onto the suffix), it would be possible, in principle, to argue that the change was motivated by phonological, rather than morphological, recoverability - since the former, but not the latter, would have been achieved. But the facts are different. The bulk of the evi-

dence which I adduced in favor of the morphological analysis consisted in observations to the effect that the conjugation system, assumed to have arisen through the restructuring of passive forms, was being regularized - a process which is complete in some Polynesian languages (e.g., Hawaiian, Tahitian) and merely well advanced in others (e.g., Maori, Samoan). I reasoned incorrectly that regularization implied restructuring. Surely regularization - i.e., the reduction of morphological ambiguity - could take place without restructuring. The evidence, therefore, does not directly support restructuring. Rather, it supports the view that recoverability is a genuine principle in the evaluation of grammars - assuming, as is reasonable, that change toward greater recoverability is in fact progressive. We cannot, on the basis of this evidence, at least, isolate phonological recoverability as linguistically distinct from morphological recoverability.

What, then, is left of the argument that the morphological analysis is correct in the case of the Polynesian passive? Before attempting to answer this question, let me introduce the Papago case (simplified in nonessential ways for the sake of space).

The points of interest can be illustrated by the third person singular possessed forms. These involve mere suffixation of [-j] to roots whose surface forms end in vowels - e.g., [mo'o] from [mo'o], [bahi] from [bahi], [gookij] from [gooki]. But when the root ends in a consonant in surface form, the suffix brings a vowel into view - e.g., [ñimaj] from [ñim], [hikaj] from [hik], [toonaj] from [toon], [ciñij] from [ciñ], [huucij] from [huuc], etc. Relevant historical events in the Piman tradition leading to Papago are the introduction of a palatalization rule, raising */t, d, n/ to [c, j, ñ] before high vowels, and the development of processes effecting the reduction or deletion of unstressed short vowels in certain environments - e.g., word-finally. Final short back vowels were deleted following any true consonant (i.e., nonlaryngeal), and final short *i was deleted following coronals. While there is evidence that deletion was chronologically prior to palatalization, modern forms show the more natural nonbleeding order to have developed at some stage (e.g., [huuñ] from *huunu). Since any of the five vowels of Piman (*i, i, u, o, a/) could occur finally, deletion gave rise to ambiguity. This ambiguity is still present in the closely related Pima of Ónavas (in Sonora), where deletion

(but not palatalization) also exists - thus, for example, in Ónavas Pima, [hik] (from *hiku) has the third singular possessed form [hikud], while [naak] (from *naaka) has [naakad]. In Papago, however, the ambiguity has been entirely eliminated (in nouns, at least) through vocalic mergers. The deleting vowels merged as follows: (1) high vowels merged to /i/ following coronals; (2) back vowels not effected by (1) above merged to /a/. These mergers result in the circumstance that the vowel appearing in the suffixed form is recoverable from the quality of the surface final consonant in the uninflected base - if the consonant is a high coronal, the vowel is [i]; otherwise, the vowel is [a]. Thus, [hik] and [naak], ambiguous in Ónavas Pima, are recoverable in Papago.

Clearly, the elimination of ambiguity here is independent of any reanalysis of inflected forms which would associate the vowel with the suffix, rather than the stem, in forms like [hikaj] and [huucij] - the vocalic mergers in no way imply such a restructuring. It is entirely consistent with the facts to assume that modern Papago simply continues synchronically the historic deletion and palatalization rules (in nonbleeding order) and that the only restructuring consists in the vocalic mergers. While a restructuring to the morphological analysis would have achieved instantaneous phonological recoverability (in this area of Papago phonology, at least), there is no evidence suggesting that the change actually happened. Morphological ambiguity is the same under either analysis - namely, zero ambiguity.

Now let us consider the Polynesian and Papago cases together. What arguments can be constructed to choose an analysis in each instance? I think that the outcome will differ in the two cases, and, moreover, that the issue will turn on 'internal' arguments (cf. Kenstowicz, 1978) of a rather traditional sort.

I will assume, since I have no evidence to the contrary, that phonological recoverability is not distinct from morphological recoverability. Instead, there is a unitary principle of (morpho-phonological) recoverability according to which the value of a grammar increases as the amount of ambiguity (in relating base and derived forms) decreases.

In the Polynesian case, the phonological and morphological analyses are equal in terms of recoverability. This equality might be formalized, for example, by designing an evaluation metric

according to which the diacritic use of a phonological segment has the same cost as does an allomorph whose distribution is not predictable from surface phonology. Clearly, then, recoverability cannot be used to decide the issue here. From this fresh starting point, we see that there is no additional cost whatsoever associated with the morphological analysis. But there is an additional cost associated with the phonological analysis - namely, the deletion rule and, assuming it to be an extra cost, the canonical disparity between underlying morpheme structure and the Polynesian syllabic canon. Given these considerations, it seems to me that the rational choice here is the morphological analysis.

In the Papago case, likewise, recoverability fails to decide the issue. Here, however, there is nothing to recommend the morphological analysis. Its choice would not eliminate the necessity for the deletion and palatalization rules, since these are independently motivated - the morphological process of perfective truncation, among other processes, exposes medial vowels to the effect of the deletion rule, and a well motivated prevocalic vowel deletion rule exposes coronals to the palatalizing effect of suffix-initial /i/. Moreover, under the morphological analysis, we must distinguish at least two types of suffixes - one having a single alternant, continuing original Piman vowel-initial (e.g., the causative-benefactive formative [-id], from *-ida), and another, continuing original consonant-initials and exhibiting synchronically three underlying forms distributed in accordance with an allomorphy rule (e.g., the modern forms deriving from Piman *-di, [-j] after vowels, [-ij] after high coronals, [-aj] elsewhere). This second type of suffix, and the allomorphy rule associated with it, are entirely a product of the morphological analysis. There is no comparable cost associated with the phonological analysis. It does not, as it would in the Polynesian case, involve a canonical disparity, since underlying morpheme structure in the phonological analysis of Papago simply corresponds to the least marked of the rich variety of syllabic patterns admitted by the Papago canon. All things considered, the phonological solution here costs no more than what is necessary in a descriptive adequate account - it is, therefore, the rational choice in this instance.

I conclude from this discussion that recoverability is a genuine principle in the evaluation of grammars. Properly construed,

however, it enters into linguistic argumentation in much the same way as do traditional cost-accounting arguments which evaluate competing analyses in terms of relative parsimony. If this is correct, then it is not surprising that recoverability may fail to decide between alternative solutions - the alternatives may, as in the two cases examined here, be equal in terms of recoverability. Beyond recoverability, there are other principles which are relevant to the evaluation of grammars, including parsimony. I very much doubt that any of these principles can be said to carry greater psychological weight than others, i.e., to be more 'real' psychologically. Our task as students of language, it seems to me, is to determine which principles are justifiable linguistically - those principles will also be justifiable psychologically, given the subject matter of linguistic science. Of course, it is legitimate in making this determination to use evidence of all sorts, and some may prove to be more helpful than others.

References

- Hale, Kenneth (1973): "Deep-surface canonical disparities in relation to analysis and change", in Current Issues in Linguistics, Volume Eleven, T.A. Sebeok (ed.), 401-458, The Hague: Mouton.
- Halle, Morris (1978): "Formal versus functional considerations in phonology", manuscript, to appear.
- Kaye, Jonathan (1975): "A functional explanation for rule ordering in phonology", in Papers from the Parasession on Functionalism, R.E. Grossman et al. (eds.), 244-252, Chicago: Chicago Linguistic Society.
- Kenstowicz, M. (1978): "Functional explanations in generative phonology", manuscript, to appear.

PSYCHOLOGICAL REALITY AND THE CONCEPT OF PHONOLOGICAL RULE

Per Linell, Dept. of Linguistics, Univ. of Uppsala, Sweden

1. Phonology is concerned with the sound patterns of various languages. In each language we use different sounds according to different rules, and the task of phonology is to define these rules. Thus, phonology is language-specific phonetics.

2. However, the usual phonological practice of most contemporary scholars in the field does not fit this description exactly. For example, in orthodox generative phonology many "low-level" language-specific phonetic regularities are not seriously considered, while many regularities which should actually belong to either lexicon or morphology are erroneously treated within phonology.

Though phonology should be concerned with speech and though speech is behavior, linguists have not studied it as behavior. Rather (some aspects of) the products of behavior have been studied in abstracto, i.e. idealized phonetic strings (words) and their interrelations have been analyzed without regard to how they are actually processed in speech production and perception and acquired by children, etc. Normally, the analysis is also crucially dependent on some kind of graphic representation. On this basis, the phonologist sets up a model of representations and rules which express connections between various idealized linguistic expressions and between properties of such expressions at various levels.

3. The problem of psychological reality in phonology concerns the relations between the representations and rules of the phonological model and the speaker-hearer's ways of storing and processing information about the structures of strings of phonetic behavior (their construction, pronunciation, recognition) and their interrelations.

4. The claims for psychological reality can be quite different in scope and content, ranging from those who assume an almost isomorphic relation between representations and rules in the phonological model and actually stored information and actual processes in speech performance, to those who see the relations as extremely indirect (the claims being therefore empirically empty). As for syntax, Fodor et al. (1974) are inclined to conclude that only the

(analysis of the) output of a standard GTG is psychologically valid. No doubt the same is true of an orthodox generative-phonological (OGPh) model (where, in practice, outputs are classical phonemic representations!). Underlying systems and derivations have no psychological reality or can be psychologically relevant only very indirectly.

5. Entities which are claimed to be psychologically valid should have plausible interpretations within (or at least be compatible with) a theory of meaningful linguistic behavior (speech). If we concentrate on phonology, i.e. on the phonetic aspects (aspects having to do with sound structure itself), what are the main problems that such a theory should be capable of solving? Perhaps the following should be mentioned:

- 5 a) How can we explain the fact that, although manifestations vary, there are many features that recur in the various manifestations of what speakers (of the same dialect) recognize as the same word form? I would propose that there is one common phonetic plan that defines the linguistic (phonological) identity of the word, a plan which specifies the linguistically relevant properties that speakers aim at realizing and which listeners tend to reinterpret into what they hear. (This is, I believe, the proper interpretation of the concept of "phonological form".)
- 5 b) How is it possible to construct phonetic plans for new forms that do not already exist as memory-stored forms? I assume that speakers may perform morphological operations which use memory-stored information to produce new phonetic plans as outputs. (These operations are naturally subordinated to the major (semantic, syntactic) intentions of the speaker's utterance construction.)
- 5 c) What is the nature of the memory-stored information used by morphological and syntactic operations?
- 5 d) How can we explain the language-specific variation in the possibilities of actually pronouncing and perceiving utterances, i.e. in the execution of utterance plans? (I assume that the phonetic aspects of an utterance plan would include at least the phonetic plans of the constituent words and a

prosodic plan of the utterance). To explain all the language-specific details of a particular utterance token, we would have to assume the existence of a fully specified articulatory plan that accounts for all the features that cannot be automatically ascribed to inherent properties of the speech apparatus. (Thus, note that the terms "phonetic plan" and "fully specified articulatory plan" are not synonymous.)

6. I have argued elsewhere (e.g. Linell 1979) that underlying morpheme-invariant forms and OGPh type derivations cannot be fruitfully incorporated into a plausible theory of meaningful phonetic behavior. Instead, there is some evidence that

- 6 a) phonetic plans (cf. 5 a) may be characterizable in terms of phonemic forms (general conditions on such forms may be stated in terms of "phonotactic rules").
- 6 b) some such phonetic plans are stored as lexical forms (stems, base forms, and some phrases) (cf. 5 b).
- 6 c) morphological operations take such memory-stored forms as inputs and produce new phonetic plans as outputs. If morphological operations are analytically split up into components, the components may correspond to morphophonological rules proper, and the whole operation will have a certain similarity to the abstract part of an OGPh derivation (except that the inputs are concrete phonetic forms rather than morphophonemic forms) (cf. (1) below).
- 6 d) the language-specific variations in normal, careful speech vs. sharpened (formal, expressive) speech and informal, casual ("fast", reduced) speech can be characterized in terms of phonological rules proper. Thus, fully specified articulatory plans may be derivable from the word-form-invariant phonetic plans (cf. 5 d).

7. In this paper I will discuss the proper interpretation of terms like rule, condition, operation, and process in phonology within a theory of the kind envisioned in §5.

Often, the discussion of the psychological reality of phonological rules is confused by the fact that several quite different concepts seem to be mixed up in most treatments.

- a) One is the (normal) interpretation of rule in the social sciences, i.e. as norm (or sometimes merely regularity) of behavior.
- b) Another one is the notion of mathematical rule, a mapping (or an instruction for the mapping) of one formally defined string of symbols onto another one.
- c) Since rules of type (b) are often described (talked about) as processes, i.e. changes of something into something else, it is sometimes tempting to interpret rules as performance processes.

The situation is further complicated in that empirically quite different sorts of regularities have often been regarded simply as "phonological rules". Thus, the putative similarities between morphophonological rules within a morphological operation like (1) and the "fast speech" rules relating different pronunciations of one and the same expression as in (2) are only superficial (and formal).

- (1) formation of noun from nonsense adjective according to the obscene-obscenity pattern:

Operand:	/rijs/
Morpholexical rule:	/rijs+it/
Trisyllabic laxing:	/risit/
Vowel shift:	/resit/

- (2) (from Donegan and Stampe, 1978) /pləntit/ plant it

Regressive nasalization:	pIə̃tɪt
Flapping:	pIə̃rɪt
Progressive nasalization:	pIə̃r̃ɪt

8. The basic concept of rule should be (7 a). Speech is a stream of phonetic behavior or phonetic events (that produce certain effects). What distinguishes speech from "mere vocalizations"

is the fact that the behavior must fulfil certain conditions of syntactic and phonological nature (and both speakers and listeners "know" this). In our model rules specify these conditions. Although behavior and actions are inherently processual, they can be looked upon either from the point of view of the processes themselves or as behavioral products. The latter is especially motivated as regards actions which are intended to produce certain effects. Thus, the act of pronouncing plant it in a certain, casual way [pɪ̃æ̃ɪ̃t̃] may be analyzed as follows: The speaker must construct a certain phonetic plan that corresponds to his communicative intentions, i.e. plant it rather than, e.g., plan it. This construction is thus subject to certain rules or conditions, which may be construed either as conditions on the behavioral operation (construction process) or on its effect (the resulting phonetic plan). The plan is then executed (realised, pronounced) in a certain way ([pɪ̃æ̃ɪ̃t̃] rather than [plæ̃ntɪ̃t̃]); the specifics of this pronunciation may be characterized as conditions on (rules for) either the pronunciation as a process or the pronunciation (or, rather, the fully specified articulatory plan) as product.

9. Note that rules concern properties of the intended behavioral products ("surface forms") (not some mystical morpheme-invariants). What these properties are must largely be determined by linguistic analysis. Thus, we cannot dispose of the traditionally linguistic (structural) analysis of language products (§2), although I would argue that (provided we are interested in psychological reality) this analysis must concern the products in relation to what we know about their production, perception, and acquisition (which means that observations of actual performance under normal and experimental conditions, slips of the tongue and the ear, child language, etc., will be of vital importance).

10. Obviously, rules as generative systems (in e.g. the OGPh fashion) need not have anything to do with conditions on actual (or potential) behavior. Indeed, the idea that behavior could be governed by generative systems seems very naive. (The various figures of figure-skating could no doubt be specified by a generative theory of figures, but who believes that the skater's behavior is produced by means of processes corresponding to such generative rules?) Thus rules are not acts or processes, but conditions on behavioral acts or on their products.

11. Behavior can be talked about at several levels of abstraction. When we talk about the morphological operations of constructing e.g. /resɪtɪ/ from /rɪ̃s/ (cf. (1)) or /fɔksɪz/ from /fɔks/ (pluralization), we are not necessarily modelling the actual behavioral process. The only thing we can say is that there is evidence that speakers can (sometimes) form "correct" ity-nouns from nonsense adjectives, that they can form plurals of English nouns, and that the respective operations are subject to certain linguistically defined conditions. That is, we can assume that speakers actually carry out morphological operations and other linguistic actions (and our models specify the linguistic content of the actions), but we cannot speculate on how these operations are neuro-physiologically implemented. Operations and actual processes lie at two different levels of description and must not be identified. Operations are defined by their intended effects, and it is conceivable that there are many ways for the neural mechanisms to achieve the goals.

It follows that rules must not be equated with behavioral processes. Not even in casual speech phonology are we entitled to conclude that rules correspond to processes. After all, conventional phonological rules state nothing but regular correspondences between idealized representations of the same or related pronunciations. (Note that I am not using 'rule' and 'process' in the way they are used in Stampean "natural phonology".)

12. I started by defining phonology as language-specific phonetics, and later I characterized rules as norms. However, this means that the phonology of a specific language would not describe or explain all the details of actual pronunciations in that language, since not all facts are conventional; some follow from biologically determined limitations. (In casual speech phonology, most regularities are language-specific variants of otherwise universal phonetic tendencies.) This is a reasonable definition of phonology, since it confines phonology to those features that must be learnt. However, we could alternatively generalize 'rules' to cover all regularities, whether conventional or biologically determined. Such a conception seems to be accepted in Stampean phonology. Thus, e.g., children's incompetence rules (i.e. Stampe's inherited processes) are clearly not social conventions. But even

such rules remain correspondence formulas; the actual phonetic processes are probably more of general continuous adjustments along scales.

13. The analysis of concepts like "psychological reality", "rule" versus "process" and "operation", etc. is necessary if the relation of phonology to phonetics is to be properly understood.

Acknowledgement

I want to thank Sven Öhman for much inspiration through the years.

References

- Donegan, P.J. and D. Stampe (1978): "The study of natural phonology", forthcoming in Current approaches to phonological theory, D. Dinnsen (ed.), Bloomington: Indiana University Press.
- Fodor, J.A., T. Bever and M. Garrett (1974): The Psychology of Language, New York etc.: McGraw-Hill.
- Linell, P. (1979) (forthcoming): Psychological Reality in Phonology, Cambridge University Press.
- Steinberg, D. (1975): "Chomsky: From formalism to realism and psychological invalidity", Glossa 9, 218-252.

EMPIRICAL INTERPRETATIONS OF PSYCHOLOGICAL REALITY

Royal Skousen, University of Texas, Austin, Texas, USA

In this paper I will discuss three requirements for a theory of language. These requirements are (1) inducibility, (2) generality, and (3) testability.

The first requirement, that of inducibility, is that linguistic descriptions must be directly derivable from the data that speakers are actually confronted with in learning the language. A linguistic description thus implies (1) a description of the relevant data and (2) a set of rules by which the linguistic description is derivable from the data. We refer to this set of rules as the rules of induction.

In order to understand this requirement of inducibility, let us consider some common violations of this requirement. For example, the order and frequency with which the data is presented to the speaker may be significant in determining the proper description of the data or in explaining how the language may change over time, so that if such information is ignored, the subsequent description may be untestable. Consider, for instance, Chomsky and Halle's statement in The Sound Pattern of English (p. 332) that "it is no doubt the case that the linguistic forms that justify our postulation of the Vowel Shift Rule in contemporary English are, in general, available to the child [?] only at a fairly late stage in his language acquisition, since in large measure these belong to a more learned stratum of vocabulary." Of course, there is no way that Chomsky and Halle's description itself can be empirically tested, since their description is based on data that, as they themselves admit, is unrepresentative of the data that children are confronted with in learning English. Children learn to speak long before they learn words as infrequent as profanity, comparative, gratitude, serenity, appellative, plenitude, divinity, derivative, conciliate, and so forth (SPE, p. 50).

Another common violation of inducibility occurs when a non-existent ordering is imposed on the data. A common method of explicating linguistic data is to first offer that data which provides direct evidence for some rule and then treat the exceptions to the rule afterwards - by adding additional rules perhaps, but without changing the original rule. Consider, for instance, Chomsky and Halle's treatment of Kasem singular and plural forms in SPE (pp. 358-364). They first give us "regular" forms like bakada and bakadi

(singular and plural for 'boy') as evidence that the singular ending is a and the plural ending is i. Then they give us the surface exceptions to this "regularity" (e.g. kambia/kambi 'cooking pot', pia/pi 'yam', buga/bwi 'river', diga/di 'room', laŋa/lə 'song', naga/nə 'leg', pia/pə 'sheep', and so on). Chomsky and Halle try to explain these forms without abandoning their original "regularity", but their explanation depends crucially upon the order of presentation of these "irregular" forms. For instance, they first argue that a plural form like kambi can be considered "regular" (that is, as /kambi+i/) if there is a phonological rule of truncation that will reduce ii to i. Having thus established that the "regular" endings are a and i and that there is a rule of vowel truncation, then a singular form like pia 'sheep' can be interpreted as /pia+a/: "Since the grammar already [!] contains the Vowel Truncation Rule, [pia] can also be derived from an underlying [piaa]." From an acquisitional point of view, Chomsky and Halle are assuming that the speaker takes care of the "regular" cases first and then the "irregular" case kambi before tackling the "irregular" case pia 'sheep' (sg.)' (which, incidentally, is "regular" on the surface). Finally, Chomsky and Halle posit a rule of metathesis for Kasem, again assuming that all rules previously posited will be maintained. The rules which Chomsky and Halle present depend upon their artificial ordering of the data. But the data is not ordered in this way for the child learning Kasem, nor does the child know in advance which of these forms are "regular" and which ones are "irregular" or "exceptional". If such a characterization of these forms is correct, then the child must discover it from random data.

Another violation is to ignore some of the data, especially those forms which the linguist knows are "incorrect": slips of the tongue, false starts, analogical creation, stuttering, dialectal variants, and so on. Yet the child does not know in advance which of the forms in the data are errors. If a child hears another child using the form goed for the past tense of go, we do not delete this from the child's data. We keep it in the data, but try to explain why the child will eventually identify goed as an incorrect past tense form. Nor should we even delete examples of stuttering from the data, since speakers can learn to imitate stuttering. Speakers also learn how to show that they have made a false start. For instance, speakers of English may use uh (but not /i/) to indicate a false start. Nor do speakers ignore dialectal variants - they learn them, even though they may not use them.

Finally, linguistic descriptions cannot be based on non-existent data. Although speakers can learn that certain items do not occur in the data, this knowledge cannot be derived from knowing in advance that these items do not occur. The determination, for example, of syntactic descriptions cannot depend upon knowing which non-occurring sentences are ungrammatical and which ones are grammatical.

The second requirement is that of generality: The rules of induction are independent of any given set of linguistic data and independent of any given regularity found in linguistic behavior. In other words, the rules of induction are universal and not taxonomic or ad-hoc. Only in this way can the explanatory goal of linguistic theory be achieved.

An excellent example of a universal rule of induction is found in Jakobson's Child Language, Aphasia, and Phonological Universals in which Jakobson proposes that "the sequence of stages of phonemic systems" found in such diverse areas as aphasia and the acquisition of languages "obeys the principle of maximal contrast and proceeds from the simple and undifferentiated to the stratified and differentiated" (p. 68). Of course, there are problems with some of Jakobson's specific claims about language acquisition, aphasia, and the phonemic systems of the languages of the world. Nonetheless, the significant contribution that Jakobson makes is that he proposes a conceptually simple and universal principle in order to explain a diversity of linguistic behavior.

In accordance with Jakobson's general principle, let us consider a principle of maximizing acoustic differences and see how it might explain the instability of certain sounds in the languages of the world. Take, for example, the case of the phoneme ü. In comparison to the phonemes i and u, ü is unstable and relatively infrequent. Children trying to learn a language that has the phoneme ü generally replace it with i or u. Historically, languages with ü frequently lose it in favor of either i or u. In the languages of the world we find phonemic systems with i-u and i-ü-u, but i-ü is relatively rare, and ü-u, as far as I am aware, is non-existent. And when i-ü does occur, it is unstable and is usually replaced historically by the more stable phonemic systems i-u and i-ü-u. Finally, when an adult speaker of an ü-less language attempts to pronounce ü, it will be pronounced as i, u, or perhaps the diphthongal iu. Now Chomsky and Halle "account for" this linguistic behavior by means of a taxonomic marking convention which simply

recapitulates the linguistic behavior formalistically (SPE, p. 405):

$$[u \text{ round}] \rightarrow [\alpha \text{ round}] / \left[\begin{array}{c} \text{aback} \\ \text{-low} \end{array} \right]$$

But a principle of maximizing acoustic differences could be used to explain this behavior. The motivation for this principle is that small acoustic differences are difficult to perceive and produce, thus shifts will occur in the direction of increasing acoustic differences. If we consider the first three formants of the vowels i, ü, and u, the maximal distinction occurs between i and u and thus the intermediate ü may be replaced by the phonetically similar i or u.

The important point in using a general principle such as this one is that it can account for the linguistic behavior of other sounds besides ü. For instance, the interdental fricatives θ and ð are also unstable and infrequent and tend to be replaced by phonetically similar sounds such as the dental fricatives s and z, the labiodental fricatives f and v, or the dental stops t and d. On the other hand, Chomsky and Halle's approach leads them to postulate a completely different marking convention in order to handle the instability of the interdental fricatives (SPE, p. 407):

$$[u \text{ strid}] \rightarrow [\alpha \text{ strid}] / \left[\begin{array}{c} \alpha \text{del rel} \\ \left\{ \begin{array}{l} [+ant] \\ [+cor] \end{array} \right\} \end{array} \right]$$

Such taxonomic rules do not explain anything; they merely formalize observed regularities. The observation of regularities is, of course, critical to the construction of a theory, but observed regularities do not make theories. Instead, regularities demand explanation in terms of general principles.

The third requirement for a theory of language is that it must be testable: A theory must have an empirical interpretation. Let us assume that we have some linguistic data for a particular language and that we apply certain rules of induction to the data and derive a description of the data. The question of utmost importance is: How can we discover if the proposed rules of induction are correct? In other words, how can we determine if the linguistic description really represents what the speaker has learned? It is not enough to simply declare that the description is psychologically real. The linguistic data is available for observation, but we cannot observe

the rules of induction that speakers are using to learn the language nor can we observe the derived linguistic descriptions. But we can observe subsequent linguistic behavior. So in order to test the rules of induction and the derived linguistic description, we need a mapping between the linguistic description and linguistic behavior. This mapping is the empirical interpretation. A theory is tested by its ability to predict the nature of linguistic behavior. Thus a theory is composed of two parts: (1) the rules of induction and (2) the empirical interpretation of descriptions. A theory without an empirical interpretation is not really a theory because it is not testable. Most so-called "theories" of language are actually rules of induction - that is, systematical methods for describing linguistic data (or deriving linguistic descriptions). Theory construction must also include the interpretation of descriptions. The empirical interpretation will predict how speakers would use the linguistic description. By comparing the predicted behavior with actual behavior we can test our theory. If a theory has an empirical interpretation (that is, if the theory is falsifiable), then we may ask if there is any evidence in favor of this theory over alternative theories and if there is any evidence against this theory. If the theory fails in some respect to correctly predict actual linguistic behavior, then the fault may lie in the rules of induction or the empirical interpretation, presuming that the linguistic data is accurately represented.

A good example of an empirical interpretation of a linguistic construct is found within those phonological theories that treat the phoneme as a psychological unit. Consider, for instance, the following possible empirical interpretations of the phoneme:

(1) Naive spellings (especially the spellings of children learning how to read and write) are based on phonemic representations. On the basis of this empirical interpretation, Read (1975, 29-78) argues that invented spellings like CHRIE for try, JRAGIN for dragon, NUBRS for numbers, LITL for little, and LADR for letter give evidence that the children's phonemic representations for these words are /čraj/, /jrægən/, /nãbrz/, /lɪtɪ/, and /lɛdr/, rather than the more common phonemic representations /traj/, /drægən/, /nambərz/, /lɪtəl/, and /lɛtər/. (These latter forms have undoubtedly been influenced by the standard orthography.) Similarly, Sapir argued (1968, 54-58) that his informants' naive spellings were also representative of their phonemic representa-

tions.

(2) Slips of the tongue are based on phonemic representations. For instance, Fromkin (1971, 33) argues that since slips of the tongue never split apart the affricates [tʃ] and [dʒ] in English, these affricates should be interpreted as single phonemes, /č/ and /ǰ/, rather than as a sequence of phonemes, /tʃ/ and /dʒ/. In contrast, actual phonemic sequences like [spr], [pɪ], [kr], [bɪ], and [fr] are frequently split apart. This difference in linguistic behavior is explained if we assume that this empirical interpretation is correct. Similarly, Stampe (1973, 35) argues that there are no archiphonemes in English because of the occurrence of [hwɪpsr̩] rather than [hwɪbsr̩] as a slip of the tongue for the word whisper. The psychological (or phonemic) representation of whisper is, say, /hwɪspr/ rather than /hwɪsbr/ or /hwɪsBr/, where B stands for a labial stop unspecified for voicing (that is, an archiphoneme). The reason then that the slip of the tongue is [hwɪpsr̩] is that slips of the tongue switch the order of phonemes, and the metathesis in this example shows that the real phonemic representation contains a voiceless, bilabial stop.

(3) Linguistic games are based on phonemic representations. This empirical interpretation serves as the basis of Sherzer's (1970) analysis of the Cuna language. The games that speakers play are characterized as simple operations on strings of phonemes, although one speaker's phonemes may be more "abstract" than another's. The problem of the English affricates can also be studied by means of linguistic games. Those "speakers" of Pig Latin who move only the first consonant of an initial consonant cluster (e.g. spin is [p^hɪnsɛj]) always move the complete affricate (e.g. chin is [ɪnčɛj] and never [sɪntɛj]), thus indicating once more that [tʃ] is to be interpreted as a single phoneme, /č/, rather than as /tʃ/.

Now let us suppose we have some rules of induction for the determination of phonemic representations and that these rules lead to the interpretation that the English affricates should be sequences of phonemes, as /tʃ/ and /dʒ/. Without any empirical interpretation of phonemic representations, there would be no way to test this description of English or the rules of induction which are used to derive this description. In order to test our theory, we must determine some empirical interpretation for our phonemic representations. If we accept these three interpretations (namely, that naive spellings, slips of the tongue, and linguistic games

are based on phonemic representations), then we can test this description of the English affricates and any set of inductive rules that would lead to such a description. We have already seen that the evidence from linguistic games and slips of the tongue imply that the affricates are unitary. In fact, children's spellings also support this conclusion, since there is no evidence for invented spellings of the form TSH for the affricate /č/ (e.g. chin is not spelled as TSHIN). In this case, all three empirical interpretations argue against the phonemic representations /tʃ/ and /dʒ/. These interpretations support each other, which is what we should expect if all three of these interpretations are correct. Now it may be that these empirical interpretations are, in fact, incorrect, but we should not reject them simply because we desire, above all else, to maintain our description of the English affricates (as /tʃ/ and /dʒ/) and the rules of induction that derive them. And even if these empirical interpretations are not correct, this does not relieve the phonologist of the responsibility to provide some empirical interpretation for his phonemic representations. In order for his theory to be testable, the linguist must determine what will count as evidence for his description and what will count against it. If the linguist can think of nothing that will disprove his theory, then he does not have a theory.

One important empirical interpretation that should hold for any theory is the principle of homogeneity: If the rules of induction do not distinguish between A and B in the linguistic description, then the behavior of A and B should be the same. Thus the rules of induction can be shown to be wrong if, in fact, A and B behave differently. The principle of homogeneity requires linguistic theory to predict linguistic behavior accurately. If a theory fails to predict an observed difference in linguistic behavior, then the theory must be revised.

A well-known case where this general principle of empirical interpretation has been used is in Kiparsky's paper "How Abstract is Phonology?". Kiparsky argued (pp. 24-25) that "contextual neutralizations are reversible, stable, and productive, whereas the alleged absolute neutralizations are irreversible, unstable, and unproductive." Now the standard generative phonology of that time did not distinguish between contextual and absolute neutralization; both were equally possible. Since linguistic behavior does distinguish between these two categories, the theory must be wrong.

Kiparsky therefore argued that the theory must include an alternation condition, which would either forbid absolute neutralization or at least make it highly improbable. In this way the linguistic theory could predict the non-homogeneous linguistic behavior.

This example suggests that the principle of homogeneity can be used to discover what sorts of information a linguistic description should have in order to predict differences in linguistic behavior. In fact, without the goal of predicting linguistic behavior, there would be no motivation for discovering the psychologically real linguistic descriptions.

References

- Berko, Jean (1958): "The child's learning of English morphology", Word 14, 150-177.
- Chomsky, Noam and Morris Halle (1968): The Sound Pattern of English, New York: Harper and Row.
- Fromkin, Victoria (1971): "The non-anomalous nature of anomalous utterances", Language 47, 27-52.
- Jakobson, Roman (1972): Child Language, Aphasia, and Phonological Universals, The Hague: Mouton.
- Kiparsky, Paul (1973): "How abstract is phonology?", Three dimensions of linguistic theory, Osamu Fujimura (ed.), 5-56. Tokyo Institute for Advanced Studies of Language, Tokyo.
- Read, Charles (1975): Children's categorization of speech sounds in English, National Council of Teachers of English, Urbana.
- Sapir, Edward (1968): "The psychological reality of phonemes", Selected Writings of Edward Sapir, David G. Mandelbaum (ed.), 46-60, Berkeley: University of California Press.
- Sherzer, Joel (1970): "Talking backwards in Cuna: The sociological reality of phonological descriptions", Southwestern Journal of Anthropology 26, 343-353.
- Stampe, David (1973): A dissertation on natural phonology, PhD dissertation, University of Chicago.

ACQUISITION OF THE PHONOLOGICAL SYSTEM OF THE MOTHER TONGUE

Summary of Moderator's Introduction

Charles A. Ferguson, Department of Linguistics, Stanford University, Stanford, CA 94305, U.S.A

The papers are fairly representative of the range of studies of phonological development now being undertaken. The period 1968-78 was one of greatly increased research in child phonology, in large part stimulated by the English translation of Kindersprache (Jakobson 1968) and the publication of experiments on speech-sound discrimination in infants (Eimas et al. 1971). A recent conference attempted to review and synthesize this research (Yeni-Komshian et al., in press).

The three areas of greatest current research effort are: neo-nate discrimination, the transition from babbling to speech (first two years of life) and the development of phonological organization (age 2-4 yrs.). The first and third are represented here directly by Kuhl and Menn, respectively, and all three are alluded to in the various papers. The phonetic/phonological development of older children is represented here by Gilbert and Hawkins, and the Hawkins paper represents the expanding field of the development of prosodic and temporal characteristics of speech.

The papers are also representative of new trends in research orientation. Earlier emphasis on innate structures and processes led to concern with (a) universal orders of acquisition of phonemes, features, and phonological oppositions, and (b) the identification of feature detectors roughly analogous to visual feature detectors. The new trend is toward emphasis on variation in the order and routes of development and on the effect of input on the child's development. The emphasis on variation is striking in Menn's paper, which classifies variation into seven types and relates these to possible developmental models, but it is evident also in Menyuk's paper, which notes that "universality is confounded by the particular data the child is confronted with." Similarly, Kuhl, who is concerned with species-wide predispositions and even predispositions shared with other species, examines the importance of "selective auditory exposure" and concludes that in infants' speech-sound category formation "their tendencies to attend to particular acoustic dimensions [are] modified by exposure to a particular language."

Another new trend is the reversal of earlier confidence that

adult models of speech processing and linguists' phonological theories are good bases for understanding child phonology. Current research tends to claim that the contribution may often go in the opposite direction, that developmental studies may help in understanding adult models and may offer a valuable corrective to phonological theory. The models offered by Menyuk and Menn, although different in approach, both illustrate this trend. Menyuk's "outside-in" model is deliberately different from linguistic segmentation and hierarchization and also suggests that current adult models may be inadequate even for adults. Menn's two-lexicon model with both non-automatic and automatic production processing has implications, not much explored by her here, for an adult model of phonology which would allow for more variation than most theories.

Knowledge of infant speech perception is increasing rapidly, as several research paradigms are followed (Morse 1974, Kuhl in press). Knowledge of "pre-linguistic" speech production is likewise increasing (cf. Dore, et al. 1976, Stark 1978, Carter 1978). Finally, both data-oriented and model-oriented studies of the development of phonological structure are increasing (e.g. Ferguson and Farwell 1975, Kiparsky and Menn 1977).

Unfortunately, however, the conceptual gap between the infant perception studies and the other studies seems to be widening. The former are perception-oriented and elaborately experimental, the latter are production-oriented and based on naturalistic observations, typically of a small number of subjects or even a single child. Ways must be found to study perception in pre-school children and to connect the neo-nate studies with studies of older children (cf. Strange and Broen, in press).

References

- Carter, A.L. (1978): "From sensori-motor vocalizations to words", in Action, gesture and symbol: the emergence of language, A. Lock (ed.), 309-349, London: Academic Press.
- Dore, J., M.B. Franklin, R.T. Miller and A.L. H. Ramer (1976): "Transitional phenomena in early language acquisition", J.Ch.Lang. 3,13-28.
- Eimas, P., E. Siqueland, P. Jusczyk and J. Vigorito (1971): "Speech perception in infants", Science 171,303-306.
- Ferguson, C.A. and C.B. Farwell (1975): "Words and sounds in early language acquisition", Lg 51,419-439.

- Jakobson, R. (1968): Child language, aphasia and phonological universals, The Hague: Mouton.
- Kiparsky, P. and L. Menn (1977): "On the acquisition of phonology", in Language learning and thought, J. Macnamara (ed.), New York: Academic Press.
- Kuhl, P.K. (in press): "The perception of speech in early infancy", in Speech and language: research and theory, N.J. Lass (ed.), New York: Academic Press.
- Morse, P.A. (1974): "Infant speech perception: a preliminary model and review of the literature", in Language perspectives-- acquisition, retardation, and intervention, R.L. Schiefelbusch and L.L. Lloyd (eds.), 19-53, Baltimore: University Park Press.
- Stark, R.E. (1978): "Features of infant sounds: the emergence of cooing", J.Ch.Lang. 5,379-390.
- Strange, W. and P.A. Broen (in press): "Perception and production of approximant consonants by three-year-olds: a first study", to appear in Yeni-Komshian et al. (eds.) Child Phonology: perception, production and deviation, New York: Academic Press.
- Yeni-Komshian, G., J.F. Kavanagh and C.A. Ferguson (eds.) (in press): Child phonology: perception, production and deviation, New York: Academic Press.

ON THE VOWEL AND ITS NATURE, BETWEEN EIGHTEEN MONTHS AND FIVE YEARS

John H. V. Gilbert, Phonetics Laboratory, Division of Audiology and Speech Sciences, University of British Columbia, Vancouver, Canada

Introduction

For a number of years, we have been interested in the development of vowels in children between eighteen months and approximately five years of age (chronological age, CA), and have conducted a number of studies which have been directed toward questions in both production and perception. In this paper I shall review them, and some other studies which relate to the title, (Gilbert a-d).

Our original curiosity about the development of vowels was motivated largely by two factors; the first being that in 1967, there was very little information relating to this particular aspect of phonological acquisition; the second being that the classic paper of Peterson and Barney (1952) tantalizingly showed marked differences in vowel formant measure between children and adults without at any place in the paper stating how old the subjects were who constituted their sample. The Peterson and Barney data showing the differences between adult males and children are illustrated in an F1/F2 plot shown in Figure 1.

Our interest in the development of vowels then developed into

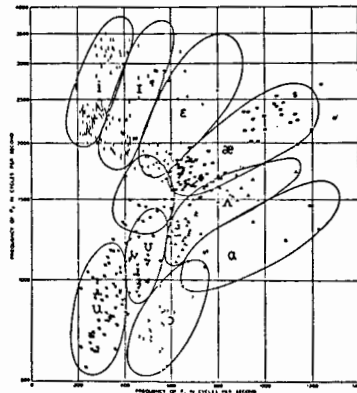


Figure 1. Frequency of second formant versus frequency of first formant for vowels spoken by men and children, which were classified unanimously by all listeners (Peterson and Barney, 1952).

three principle questions: the first was whether it was possible to

accurately trace the development of vowel sounds from around eighteen months to their adult values; in this we were superceded by the excellent work of Eguchi and Hirsch (1969) whose formant measures for vowels over time are shown in Figure 2.

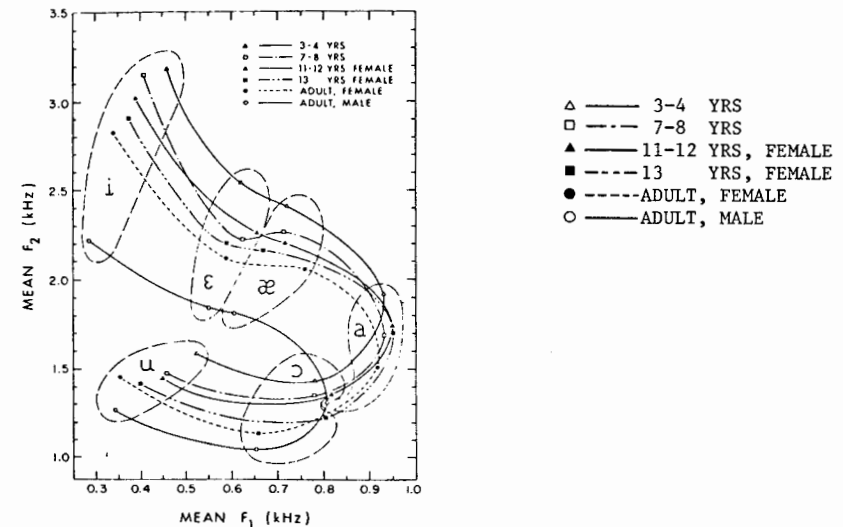


Figure 2. Mean formant frequencies for combined age groups as shown in the key. Each point represents the combination of Formant 1 and Formant 2 for each of the six vowels. The different symbols together with the lines that join them represent the different ages. The broken circles are drawn around all points for a given vowel. (Eguchi and Hirsh, 1969).

There were, however, subsidiary questions relating to the problem of the ontogeny of vowels, in particular, whether children of the same chronological age but at different stages of physiological development, would demonstrate differences in vowel formant frequencies because of their differences in growth. I will report this information later.

The second principle question related to formant measures of vowels produced by groups of children who were measureably different in their linguistic development, since a great deal of space in the phonological literature has been (and continues to be) devoted to a discussion of how and in what sequence consonant sounds emerge. We felt that children at different stages in the acquisition process might give us some information relating to this question, at least for vowels.

A third, and last principle question, concerned the manner in which vowel sounds are perceived by children when these sounds are produced both by themselves and adults. Since vowels are perhaps more easily acoustically measured than consonants in the output of children, and since there appears to be more listener agreement on their character, we considered this line of investigation was one worth following.

Studies of Vowels

Because of their clear separation in the vowel quadrilateral, our energies were directed chiefly to an examination of four vowels: /i/ as in "heed", /æ/ as in "had", /ɒ/ as in "hod", and /u/ as in "who'd", produced by both children and adults, usually in an h-d environment. Other studies reported in the literature have examined a wider set than this. The choice of these vowels, however, allowed us to compare our results with results from numerous studies conducted with adults, and in retrospect, to consider some issues, e.g. individual variation, as they apply to the emergence of vowels during acquisition. Bearing in mind the problems of holding "mechanical" (i.e. child) parameter constant, and the difficulties of minimizing measurement variation (see Kent, 1976, 1978 for details on acoustic analyses of children's vowels), we hoped to view the vowel system "settling down" across chronological age.

In an early paper, Okamura (1966) measured five vowels spoken by 475 Japanese children and demonstrated that the formant frequency construction of these vowels was quite different between children and adults. A copy of his centre formant frequency measurements is shown in Figure 3.

It will be seen that for all of these vowel sounds, the formant frequency measurements appear to plateau around seven years of age. When we came to compare our own data for four-year-old English speaking children with that of Okamura, we found a fair measure of agreement for formant two. Our measurements are shown in Table 1.

Interestingly, the use of duration of vowels in emerging phonology appears, at least in one report (Di Simoni, 1974), to follow this development trend; by age six, durational differences between vowels becoming stabilized in children's speech. This issue is, however, confused and the reader is referred to a comprehensive account of factors in Greenlee (1978).

As mentioned earlier, one of our interests was to determine

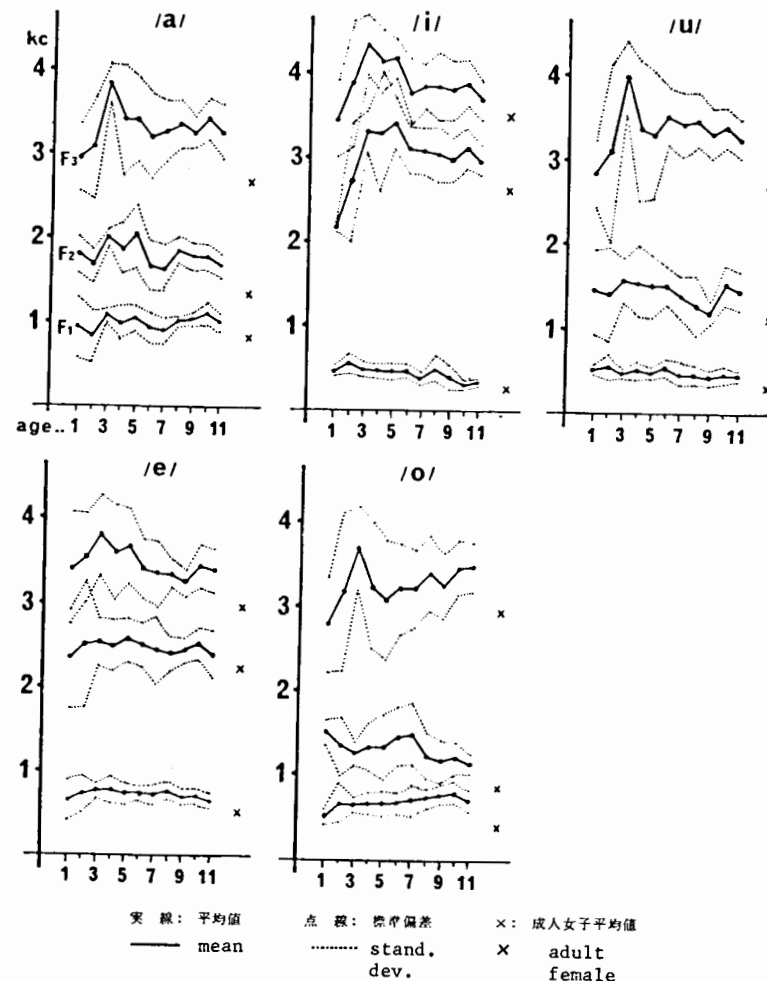


Figure 3. Formants and their standard deviations (Okamura, 1966)

wether differences in physiological age (whilst holding chronological age constant) would, in fact, change the acoustic characteristics of children's vowel sounds. It did not appear sensible to group children by CA for the purposes of examining vowel development if, in fact, their physiological ages were markedly dissimilar. The motivation for this observation was the assumption that a difference in physiological age would mean a difference in vocal

tract length and therefore a difference in characteristic vocal tract resonances.

Table 1. Vowel productions: means and standard deviations, in hertz for F1 and F2 measurements of control and experimental groups (Gilbert, 1970).

Vowel		F1		F2	
		Mean	Stand. dev.	Mean	Stand. dev.
/i/	Control	442	107	2510	99
	Experimental	555	149	2613	67
/æ/	Control	917	183	1710	251
	Experimental	859	130	1631	122
/a/	Control	693	112	1246	157
	Experimental	727	113	1216	299
/u/	Control	539	166	1255	202
	Experimental	533	115	1336	207

We found that both F1 and F2 naturally show a tendency to drop with an increase in chronological age from fourteen to eightyfour months, and that when subjects were reassigned to groups by Bone Age (BA) (Harrison, et al. 1964) groupings, (BA being the physiological measure which we used), the same pattern emerged. We found no statistical difference between Ca and BA on our formant measures; we thus concluded that grouping children by a measure of physiological maturity (rather than CA) does not in the final analysis alter results.

In retrospect, I am not sure that this was an appropriate conclusion to draw, based on the way in which we assigned children to BA groupings. I suspect that it would have been more appropriate to have taken both BA plus skeletal size, i.e. height and weight, and then compared them with children of similar CA and intelligence. We know from the work of Negus (1949) that the larynx develops most rapidly between 3;0 and 5;0 CA and then increases in size to maturity very slowly. This point should have been taken into account. I am still not convinced that we have solved the physiological age problem in our deliberations.

From a consideration of physiological age we then moved to a slightly different view of the process, that is, would children at different levels of linguistic development, but at the same CA exhibit any significant differences in vowel production. Given that children are the same height and weight our assumption would be

one of no difference, since we would expect that whatever emerged from the vocal tract would be of the same order, regardless of whether or not each child's linguistic abilities were different. We recorded children at 4;0 CA divided into groups on the basis of normal and late language usage. Although there were no differences between these groups in terms of mean formant two measurements, when we played tokens produced by the late language users to adult listeners for identification, the adults were definitely confused in their perception, a result which we had certainly not anticipated.

We interpreted this discrepancy as an indication that children who are at a less mature stage of linguistic development are doing "something" to the vowels which cannot be accounted for on an acoustic basis. A thorough examination of the acoustical similarities and dissimilarities between normally developing children and language delayed children is necessary before we can make any further judgements. It may well be that the dynamic acoustic information distributed over the temporal course of the syllable, is affecting listener judgement differently in each case.

The last question to which we addressed ourselves involved the perception of vowels. In 1967, Menyuk reported an experiment in which she showed that the phoneme boundaries for a set of vowels in consonantal context were the same for six children between 5;0 and 10;0 as they were for adult listeners. We found in our experiments that children at 4;0 have no difficulty in discriminating four broadly spaced vowel tokens spoken by themselves and by adults, when these are presented to them in an h-d context. We also found that, when children at this age are asked to produce vowel sounds in an h-d context (in response to the same vowels in h-d context spoken by adults and child speakers other than themselves), there is virtually no difference between F0 and F2 in their tokens, and the tokens of the speakers whom they are imitating.

Lieberman (1978) and his associates at Brown University have data which shows a gradual and consistent improvement in the children's productions of vowels of English from the early stages of babbling through to 3;0; an age at which the children are using meaningful sentences and conversing with the experimenters. Lieberman's data is very robust, and certainly corroborates our own notions about vowel development.

Conclusion

The question of the acoustical development of vowel sounds appears to be reasonably well answered by now. That is, one sees an increasing trend over the first six years towards the adult form in terms of fundamental frequency, F2 and F3. The plateauing between 6;0 and 8;0 is undoubtedly related to the fact that the vocal tract at this time is approaching its adult measurement. The question of the child's perception of vowel sounds appears more equivocal. Since perception will have to be accounted for by correct usage in production, we will need further experiments of the kind recently reported by Greenlee (1978) before any definitive statements can be made. The same reservation is also true for adult listeners' perceptions of child talkers. There appears to be minimal evidence that children attempt to mimic the acoustic characteristics of the adult speech that they hear, although we do know from Garnica (1974) that at least the mother is adjusting the acoustic characteristics of her utterances to the child. Why is it then that children's vowel utterances are so clearly delineable at a relatively early age? As discussed by Verbrugge et al. (1976) normalization does not appear to be a satisfactory answer.

References

- DiSimoni, F.G. (1974a): "Evidence for a theory of speech productions based on observations of the speech of children", JASA 56, 1919-1921.
- DiSimoni, F.G. (1974b): "The effect of vowel environment on the duration of consonants in the speech of three-, six-, and nine-year-old children", JASA 55, 360-361.
- DiSimoni, F.G. (1974c): "Influence of consonant environment on the duration of vowels in the speech of three-, six-, and nine-year-old children", JASA 55, 362-363.
- Eguchi, S. and I.J. Hirsh (1969): "Development of Speech Sounds in Children", Acta Oto-Laryngologica 257, 7-51.
- Garnica, O. (1974): Unpublished Ph.D. Diss., Stanford University.
- Gilbert, J.H.V. (1970): "Formant Concentration Positions in the Speech of Children at Two Levels of Linguistic Development", J. Acoust. Soc. Amer. 6,2, 1404-1406.
- Gilbert, J.H.V. (1970): "Vowel Production and Identification by Normal and Language Delayed Children", J. Exper. Child Psych., 9, 12-19.
- Gilbert, J.H.V. (1973): "Acoustical Features of Childrens' Vowel Sounds: Development by Chronological Age Versus Bone Age", Language and Speech 16,3, 218-223.
- Gilbert, J.H.V. (1977): "The Identification of Four Vowels by Children 2½ to 3 Years Chronological Age as an Indicator of Perceptual Processing", In, Segalowitz, S. and F. Gruber (eds.), Language Development and Neurolinguistic Theory, New York: Academic Press, Chapter 19.
- Gilbert, J.H.V. and V.J. Wyman (1975): "Discrimination Learning of Nasalized and Non-Nasalized Vowels by Five-, Six-, and Seven-Year-Old Children", Phonetica 31, 65-80.
- Greenlee, M. (1978): Unpublished Ph.D. Diss., University of California, Berkeley.
- Harrison, G.A., J.S. Weiner, J.M. Tanner and N.A. Barnicot (1964): Human Biology, Oxford: The University Press.
- Kasuya, H., H. Suzuki, and K. Kido (1968): "Changes in Pitch and first three formant frequencies of five Japanese vowels with age and sex of speakers", Research Institute of Electrical Communication, Tokuki University, 344-346.
- Kent, R.D. (1976): "Anatomical and neuromuscular maturation for the speech mechanisms: Evidence from acoustic studies", JSHR 19, 421-447.
- Kent, R.D. (1978): "Imitation of synthesized vowels by pre-school children", J. Acoust. Soc. Amer. 63,4, 1193-1198.
- Lieberman, P. (1978): "On the Development of Vowel Production in Young Children", Paper presented at "Child Phonology, Perception, Production and Deviation", Bethesda, Md. May 28-31.
- Menyuk, P. (1967): "Children's Perception of a Set of Vowels", QPR 84, M.I.T., 254-313.
- Negus, V.E. (1949): "The Comparative Anatomy and Physiology of the Larynx", N.Y.: Hafner.
- Okamura, M. (1966): "Acoustical Studies on the Japanese Vowels in Children", Japanese J. Otol. 69,6, 1198-1214.
- Peterson, G.E. and H.L. Barney (1952): "Control methods used in a study of the vowels", JASA 32, 175-184.
- Verbrugge, R.R., W. Strange, D.P. Shankweiler, and T.R. Edman (1976): "What information enables a listener to map a talker's vowel space", J. Acoust. Soc. Amer. 60,1, 198-212.

THE CONTROL OF TIMING IN CHILDREN'S SPEECH

Sarah Hawkins, Dental Research Center, University of North Carolina, Chapel Hill, NC 27514, U.S.A.

It is generally acknowledged that temporal and prosodic variables significantly affect speech intelligibility. For example, adult listeners derive considerable information about the syntax and stress pattern of sentences, when segmental cues are either distorted by spectral rotation or absent because the sentence is hummed. The role of prosody in defining syntactic boundaries has been demonstrated with stylised synthetic intonation contours and with prosody pitted against syntax in cross-spliced sentences. The duration alone of both phonemes and larger units can be crucial to speech intelligibility. Additionally, adults appear to be particularly sensitive to the rhythmic onset of stressed syllables, both when listening to speech and when tapping to the rhythm of their own speech. The listener appears to anticipate when stresses will occur and focuses attention at these times. (Documentation of the above points may be found in papers for the other Symposia of this Congress and in Cohen and Nooteboom (1975).) This integrative and predictive role of prosodic cues figures prominently in recent models of speech perception. For example, Pisoni and Sawusch (1975) suggest prosodic cues may form an interface between low-level segmental information and higher levels of syntax and meaning. Martin (1972) has elaborated the notion of the predictive role of rhythm in speech perception, pointing out that efficient perceptual strategies such as attention-cycling between input and output can be facilitated when the signal need not be monitored continuously.

What relevance have these observations about adult speech to children's perception and production of speech? Although the adult listener may be assumed to attend only minimally to much of the acoustic signal, this cannot be assumed for the child. The young child lacks the linguistic and nonlinguistic experience that would allow him/her to "fill in" a large proportion of the message on the basis of knowledge shared with the speaker. Our sparse knowledge of children's perceptual abilities supports this distinction between adults' and children's perception of speech. For

example, although infants of less than 16 weeks can discriminate between stimuli differing in some durational aspects, such as VOT (e.g. Eimas et al., 1971) and syllable duration (Spring and Dale, 1977), children as old as 4-6 years do not necessarily use these durational cues in the same way as adults (Zlatin and Koenigsnecht, 1975; Simon and Fourcin, 1978; Higgs and Hodson, 1978). We know more about children's speech production than their perception; the last 5 or 6 years have provided data on children's timing in phrases, words, syllables (Hawkins and Allen, 1978), segments, and subsegmentals such as VOT (for additional references see below). Many of these studies have demonstrated that while some aspects of children's speech timing resemble those of adult speech from quite an early age, (about 2-4 years), other aspects do not mature until much later (up to about 9-11 years).

The question becomes when and why there are differences between adults and children. Do timing rules appear in children's speech simply as a consequence of increasing neuromuscular coordination, and only gradually come to serve a perceptual function? Or does the child learn the perceptual function of such cues and attempt to produce them in his/her own speech? In the latter case the age when adult timing relationships appear would depend partly on neuromuscular abilities and partly upon the age when their perceptual function is recognised.

This paper discusses ways in which we might distinguish between the above possibilities, using data from children's speech production. The aim is to provide a conceptual framework that will be useful in thinking about children's timing control, as a first step towards formulating a theory of the development of speech timing.

I begin from the position that the child's perception of speech neither is essentially mature before s/he begins to speak (cf. Smith, 1973), nor develops concurrently with production (cf. Waterson, 1970, 1971a,b). Rather, I shall assume that while the young speaker perceives some parts of the speech signal quite maturely, s/he perceives other parts only as unanalysed 'noise'. This view is similar to that of Ingram (1974), except that I assume that the position of the 'noise' may not be fixed in the signal in a segment-by-segment manner. The approach is polysystemic: the child's systems of perception and production both may

be described in terms of quasi-independent subsystems, any or all of which may be in a state of considerable flux at a given time.

I assume also that processes manifested in the child's speech will appear in adult speech and most (but perhaps not all) processes of adult speech will appear in child speech. What distinguishes the two is the domain of influence of each process. Thus in adult speech we may find evidence for both hierarchically integrated "comb" models and sequentially ordered "chain" models of timing (Bernstein, 1967; Kozhevnikov and Chistovich, 1965; Ohala, 1975). Phonemes may be integrated into syllables, for example, consistent with the "comb" model, while the timing of successive higher units may be relatively independent of each other, consistent with a "chain" model. In the young child's speech, such integration implying a "comb" model may not be evident at the syllabic level, but similar integration may occur with less complex units. This reasoning suggests that the child's task in learning to speak fluently is not so much that of learning new routines as of applying similar routines to increasingly more complex domains, thereby integrating the elements of these domains into functional units (c.f. Turvey's (1977) action plans). During this learning, the role played by different processes may change. Auditory feedback in children's speech, for example, appears to have a qualitatively different effect than in adults' speech (e.g. Fry, 1966; MacKay, 1967).

Bearing the above assumptions in mind, let us consider some of the different factors that are likely to affect the speech learning process, together with examples of evidence for their existence within speech timing. Three such factors will be distinguished of which only the second and third are mutually exclusive: (1) processes common to all motor skill development; (2) temporal distinctions that serve as primary perceptual cues; and (3) temporal regularities that do not function as primary perceptual cues.

Processes common to all motor skill development will apply to all aspects of speech development; examples are slower and more variable performance. These phenomena have been demonstrated for speech in many studies and at many levels of analysis from the phrase to the segment (e.g. Eguchi and Hirsh, 1969; DiSimoni, 1974a,b; Tingley and Allen, 1975; Kent and Forner, 1977; Keating and Kubaska, 1978; Smith, 1978; Hawkins, in press). It could be

argued that slower durations occur because the child possesses an articulation-dominant system, whereas the (English-speaking) adult uses a timing-dominant system (cf. Ohala, 1970). Nevertheless, the basis of such articulation-dominance is plausibly immature neuromuscular coordination, requiring more time to achieve adequate articulatory targets. Where a phonological length distinction occurs, the child seems to learn to shorten rather than lengthen articulatory units to produce it. This has been suggested for example for the effect of position-in-utterance on word duration (Keating and Kubaska, 1978), for the development of unstressed syllable production (Allen and Hawkins, in press), and in older children for the effect of clustering on consonant duration (Gilbert and Purves, 1977; Hawkins, in press).

Even in such apparently simple cases as longer duration, a single effect may have more than one underlying cause. For example, measuring /b, d, t/ durations in simple environments, Smith (1978) observed that /t/ was 40% longer in the speech of 2 and 4 year olds than might be expected just on the basis of estimates of the degree of durational increase due to neuromotor immaturity. Smith suggested 3 possible causes for this, any or all of which may have contributed to the observed effect: (i) an effort to increase perceptual differences between /t/ and /d/, (ii) greater complexity of the tongue tip innervation required for /t/ than /d/, and (iii) greater complexity of laryngeal adjustments for voiceless stops over voiced ones.

Temporal distinctions that serve as primary perceptual cues are likely to be detected by the child relatively early as long as they do not signal, for example, syntactic or semantic distinctions beyond the child's comprehension. Hence they should appear in the child's speech in an order reflecting the complexity of the neuromotor coordination required. I shall discuss two examples: phonemically conditioned vowel duration, and voice onset time (VOT) in stops.

Vowel duration functions in English as a primary perceptual cue to the voicing of following consonants, with longer vowels preceding voiced consonants. There is some evidence that this is a distinction that occurs naturally and has been exaggerated in some languages, such as English (Lisker, 1974). Such evidence would suggest that the child might learn to produce the mature

voiced-voiceless ratios relatively early. Naeser (1970) found they were present by 21 months of age and in fact preceded control of the voicing feature that governs the distinction in adult speech.

The development of VOT control in stops nicely illustrates the following points: (i) perception and production do not always develop hand in hand, and (ii) a phonemic distinction that may be legitimately regarded as lying along a single phonological dimension should not necessarily be treated as two extremes of a unitary process in a theory of speech development. The development of the voicing contrast in English has been studied longitudinally (e.g. Kewley-Port and Preston, 1974; Macken and Barton, 1978) and cross-sectionally (e.g. Menyuk and Klatt, 1975; Zlatin and Koenigsnecht, 1976; Gilbert, 1977). It has been consistently found that children make a distinction between short-lag (voiced) and long-lag (voiceless) stops fairly early (around 2 years), but at this stage only the short-lag distribution resembles the adult form. It is not until much later (around 6 years or more) that the long-lag VOT distribution resembles that of the adults. The difference in age of mastery of the two VOT categories is commonly accepted as being due to differences in the neuromuscular coordination required: short-lag stops allow considerable variability in the coordination of laryngeal and oral activity, whereas long-lag stops require rather precise coordination.

Temporal regularities that do not function as primary perceptual cues, especially those that appear to provide no perceptual information, would be expected to be acquired as the child's action plans (articulatory programs) become more sophisticated. An example is the reduction of the duration of consonants in clusters. Although many of the durational differences between clustered and unclustered consonants are perceptible, they may not serve a perceptual function (Klatt, 1976). The age when children produce these durational modifications varies according to the type of cluster, but the last ones are probably not mastered until 9-11 years (Gilbert and Purves, 1977; Hawkins, in press). This is later than many of the temporal phenomena discussed above. Hawkins (in press, in preparation) discusses evidence from these data suggesting that many clusters undergo considerable reorganisation as complex units, rather than there being simply durational

reduction of each segment. Intermediate stages may involve uneven rates of development and durational changes in the opposite direction from that finally required. These observations, together with the late attainment of mature durational relationships, are consistent with the development of increasingly sophisticated motor action plans by the integration of subroutines.

This paper has discussed some of the considerations that should be included in a theory of the development of speech timing within a polysystemic, parallel processing approach. It suggests that adults and children will differ in speech production processes not so much in the nature of those processes as in their relative importance and their domain of influence. The child's 'system' cannot be regarded as static at any time, but rather as reflecting the effects of several continually changing systems that replace each other during development. Changes in one subsystem may affect others, producing either progression or temporary regression. Furthermore, a given phenomenon observed in development may have several causes, whose effects may all work in the same direction or in conflicting directions. Consistent with these points, the development of speech timing may be usefully considered in the contexts of maturing neuromuscular coordination and the perceptual cueing function of timing rules. Neuromuscular immaturity influences patterns of production both across-the-board and in specific contexts. It is reasonable to expect timing rules that are primary perceptual cues to be implemented earlier than those that are not, assuming both that the degree of neuromuscular complexity is constant and that the child perceives the distinction as linguistically relevant.

Bibliography

- Allen, G. D. and S. Hawkins (in press): "Trochaic rhythm in children's speech", in Current Issues in the Phonetic Sciences, H. Hollien and P. Hollien (eds.).
- Bernstein, N. A. (1967): The Coordination and Regulation of Movements, Oxford: Pergamon Press.
- Cohen, A. and S. G. Nootboom (1975): Structure and Process in Speech Perception, Berlin: Springer-Verlag.
- DiSimoni, F. G. (1974a): "Effect of vowel environment on the duration of consonants in the speech of 3-, 6-, and 9-year-old children", JASA 55, 360-361.
- DiSimoni, F. G. (1974b): "Influence of consonant environment on duration of vowels in the speech of 3-, 6-, and 9-year-old children", JASA 55, 362-363.

- Eguchi, S. and I. J. Hirsh (1969): "Development of speech sounds in children", AcOtolaryng 257, 1-51.
- Eimas, P., E. Siqueland, P. Jusczyk and J. Vigorito (1971): "Speech perception in infants", Science 171, 303-306.
- Fry, D. B. (1966): "The development of the phonological system in the normal and the deaf child", in The Genesis of Language, F. Smith and G. A. Miller (eds.), 187-206, Cambridge, Mass.: MIT Press.
- Gilbert, J. H. V. (1977): "A voice onset time analysis of apical stop production in three-year-olds", J. Ch. Lang. 4, 103-110.
- Gilbert, J. H. V. and B. A. Purves (1977): "Temporal constraints on consonant clusters in child speech production", J. Ch. Lang. 4, 417-432.
- Hawkins, S. (1973): "Temporal coordination of consonants in the speech of children: Preliminary data", JPh 1, 181-217.
- Hawkins, S. (in press): "Temporal coordination of consonants in the speech of children: Further data", JPh.
- Hawkins, S. (in preparation): "Processes in the development of speech timing control".
- Hawkins, S. and G. D. Allen (1978): "Acoustic-phonetic features of stressed syllables in the speech of 3-year-olds", JASA 63, Suppl. 1, S56.
- Higgs, J. W. and B. W. Hodson (1978): "Phonological perception of word-final obstruent cognates", JPH 6, 25-35.
- Ingram, D. (1974): "Phonological rules in young children", J. Ch. Lang. 1, 49-64.
- Keating, P. and C. Kubaska (1978): "Variation in the duration of words", JASA 63, Suppl. 1, V10.
- Kent, R. D. and L. L. Forner (1977): "A developmental study of speech production: Data on vowel imitation and sentence repetition", JASA 62, Suppl. 1, 0015.
- Kewley-Port, D. and M. Preston (1974): "Early apical stop production: A voice onset time analysis", JPh 2, 195-210.
- Klatt, D. H. (1976): "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence", JASA 59, 1208-1221.
- Kozhevnikov, V. A. and L. A. Chistovich (1965): "Speech: Articulation and perception", Joint Publications Research Service, Washington, D. C.: 30, 543.
- Lisker, L. (1974): "On 'explaining' vowel duration variation", Haskins Labs.: Status Report on Speech Research SR-37/38, 225-232.
- MacKay, D. G. (1967): "Metamorphosis of critical interval: Age-linked changes in the delay in auditory feedback that produces maximal disruption of speech", JASA 43, 811-821.
- Macken, M. A. and D. Barton (1977): "A longitudinal study of the acquisition of the voicing contrast in American-English word-initial stops, as measured by voice-onset time", Stanford University: Papers and Reports on Child Language Development 14, 74-120.
- Martin, J. G. (1972): "Rhythmic (hierarchical) versus serial structure in speech and other behavior", Psych. Rev. 79, 487-509.
- Menyuk, P. and M. Klatt (1975): "Voice onset time in consonant cluster production by children and adults", J. Ch. Lang. 2, 223-231.
- Naeser, M. A. (1970): "The American child's acquisition of differential vowel duration", Madison, Wisconsin: Technical Report 144.
- Ohala, J. J. (1970): "Aspects of the control and production of speech", UCLA Working Papers in Phonetics, 15.
- Ohala, J. J. (1975): "The temporal regulation of speech", in Auditory Analysis and Perception of Speech, G. Fant and M. A. A. Tatham (eds.), 431-453, London: Academic Press.
- Pisoni, D. B. and J. R. Sawusch (1975): "Some stages of processing in speech perception", in Structure and Process in Speech Perception, A. Cohen and S. G. Neebboom (eds.), 16-35, Berlin: Springer-Verlag.
- Simon, C. and A. J. Fourcin (1978): "Cross-language study of speech-pattern learning", JASA 63, 925-935.
- Smith, B. (1978): "Temporal aspects of English speech production: A developmental perspective", JPh. 6, 37-67.
- Smith, N. V. (1973): The Acquisition of Phonology: A Case Study, Cambridge, England: The University Press.
- Spring, D. R. and P. S. Dale (1977): "Discrimination of linguistic stress in early infancy", JSHR 20, 224-232.
- Tingley, B. M. and G. D. Allen (1975): "Development of speech timing control in children", Ch. Dev. 46, 186-194.
- Turvey, M. T. (1977): "Preliminaries to a theory of action with reference to vision", in Perceiving, Acting, and Knowing: Toward an Ecological Theory, R. Shaw and J. Bransford (eds.), Hillsdale, N. J.: Lawrence Erlbaum Associates.
- Waterson, N. (1970): "Some speech forms of an English child: A phonological study", Transactions of the Philological Society, London: 1-24.
- Waterson, N. (1971a): "Child phonology: A prosodic view", JL 7, 179-211.
- Waterson, N. (1971b): "Child phonology: A comparative study", Transactions of the Philological Society, London: 34-50.
- Zlatin, M. A. and R. A. Koenigsknecht (1975): "Development of the voicing contrast: Perception of stop consonants", JSHR 18, 541-553.
- Zlatin, M. A. and R. A. Koenigsknecht (1976): "Development of the voicing contrast: A comparison of voice onset time in stop perception and production", JSHR 19, 93-111.

CROSS-LINGUISTIC EVIDENCE ON THE EXTENT AND LIMIT OF INDIVIDUAL VARIATION IN PHONOLOGICAL DEVELOPMENT

David Ingram, University of British Columbia, Vancouver, B.C., Canada

Let me begin by describing four types of variation that could occur during children's phonological development:

1. intra-child variation: the production of different phonetic forms for the same word, or the inconsistent use of a particular phonological process across words by an individual child;
2. inter-child variation: the production of different phonetic forms or different phonological processes by different children at a comparable stage of development;
3. intra-language variation: the occurrence of different developmental stages or patterns by children learning the same language;
4. inter-language variation: the existence of phonological processes for all children learning a particular sound pattern in one language distinct from those used by all children learning a similar pattern in another language.

We need to determine, first, do all four possibilities actually occur in development? and second, what is the extent and limits of each? Here I will briefly demonstrate that the first three exist, and then comment on what it means to look for inter-language variation.

Intra-child variation, or the use of varying phonetic forms by the same child has been noted for a long time in phonological diaries. Recently, investigators have attempted to document and explain this occurrence. A succinct and plausible description of what may be occurring is expressed in Klein (1977, 159): "It appears then, that the amount of variation in a child's productions may be a function of the type and variety of processes a child has available and applies in modifying words as he/she attempts to say them".

The existence of inter-child variation is also well-documented. Children learning the same words and sounds at comparable periods of development will often show varying ways to produce them, seemingly due to preferences for particular sounds or syllabic shapes, e.g. Priestley (1977). Klein (1977) has dealt with one such pattern at length, that of a preference for reducing syllables versus

one for syllable expansion as in reduplication. Ferguson, in several papers, has referred to such alternatives as individual strategies and suggests that the extent of variation may be quite great. For example, Ferguson and Farwell (1975, 437) in a study on the phonological development of three children during the first 50 words conclude: "each of the three children is exhibiting a unique path of development, with its individual strategies and preferences and an idiosyncratic lexicon".

The occurrence of inter-child variation at any stage implies the existence of intra-language variation, i.e. different developmental stages or paths. Elsewhere, however, (Ingram 1974b, 1978) I have argued that one needs to be cautious about claims of widespread variation and alternative stages. Even though variation in the production of specific words and the use of phonological processes may occur, a broader view may indicate that this variation is simply part of a more general pattern. Regarding phonological processes, I suggest that children may follow them in varying degrees (Ingram, 1974b). For variations that result from preference for certain sounds, it may also be that more similar or general patterns are discernible once sound classes are observed (Ingram, 1978).

So far, little research has been done on inter-language variation, with the prevailing position being that most cross-linguistic data will be similar to intra-language findings. In my earlier study of general phonological processes (Ingram, 1974a), I observed that these tended to occur with children acquiring different languages, a point also made in regard to the less obvious process of fronting (Ingram, 1974b). While these claims deal with aspects of development shared by children, similar ones can be found for variations between children. In Ingram et al. (to appear), we studied the acquisition of English word-initial fricatives and affricates across 73 children, determining their individual preferences. Most generally, one could say that when preferences occurred, they were for either labial, alveolar, or palatal productions. This three-way possibility was also observed in Ingram (in press) for three French children, Elie-Paul (Vinson, 1915), Fernande (Roussey, 1899-1900), and Suzanne (Deville, 1890-91) as found in their substitution summarized in table 1.

Table 1

Substitution patterns for French fricatives by three French-learning children

Adult Sound	Substitutions		
	Elie-Paul (1;11)	Fernande (2;4)	Suzanne (1;11)
/f/	-	s	f
/v/	-	s	v
/s/	ʃ	s	s(t)
/z/	ʃ	-	z
/ʃ/	ʃ	s	s
/ʒ/	ʃ	s	ʒ

Is there, then, inter-language variation, i.e. the existence of distinct phonological processes for learners of one language that do not occur for learners of another language? To begin, there are two simplistic extremes that reduce the issue to a trivial one. On the one side, we can say that children are all genetically prepared to learn language, and consequently all have the same processing mechanisms available. In this case, no inter-language variation is possible, for it may simply be that the proper circumstances for a particular process do not apply. For example, a child will not show simplification of consonant clusters in learning a language that does not have them. The other situation is to say that all languages are phonologically different, so that of course children will show differences across languages because there are distinct phonetic inventories and phonological patterns to be learned.

There is, however, a middle ground between these two where variation may be viewed in a non-trivial fashion. This is where there are cases when two languages appear to present children with similar phonological patterns, but children do not deal with them in the same way. A closer analysis should provide insight into the multiple conditions that occur in a particular language which lead to different learning patterns across languages.

Let me provide two examples of such patterns. In English, there is a common process of velar assimilation where an alveolar consonant will assimilate to a following stop if it is velar, e.g.

Jennika 1;7 duck [gʌk]; 2;2 tickle [gigu]. This occurs in both CVCs where the sounds are within a syllable, and in CVCV words where the assimilating sounds cross a syllable boundary. I have examined extensive data from the three French children mentioned above and have not found instances of this process. Possible explanations are that potential instances are rare, and that the different timing pattern of French inhibits its occurrence. Nonetheless, one can locate places where it could occur, given our current understanding of the process.

A second example concerns a process that all three French children show quite widely, but which is not found in English learning children with any frequency. This is the process of consonant denasalization where a nasal consonant will denasalize in harmony with a nonnasal obstruent, e.g. Fernande mange 'eat' [baʃ]; menton 'chin' [ba:to:]; marcher 'walk' [base]. Table 2 presents data indicating how dominant the pattern is for the three French children under discussion.

Table 2

Proportion of occurrence of denasalization for three French children at varying ages

age	Suzanne	age	Fernande	age	Elie-Paul
1;0-1;7	.00 (0/1)	1;4-1;9	1.00 (5/5)	2;1	.50 (2/4)
1;8	.67 (4/6)	1;10-2;0	.83 (5/6)	2;2-2;5	1.00 (7/7)
1;9	.67 (6/9)	2;1-2;3	1.00 (8/8)	2;6-3;0	.25 (2/8)
1;10	.50 (6/12)	2;4-2;7	.25 (1/5)		
1;11	.43 (6/14)	2;8-2;10	.00 (0/7)		
2;0	.19 (3/16)				

Like the previous example, one can cite some factors that might contribute to its nonoccurrence in English, but potential cases do arise.

Data like these suggest that inter-language variation occurs, and that we need to seek more instances of it. Once more are found, they should show that phonological acquisition is the complex interaction of several conditions in the adult language, and that phonological processes will need to be described in much more detail

than exists to date. Also, they indicate that the study of acquisition in only one language may yield a restrictive, and possibly misleading, view of the language learning process.

References

- Deville, G. (1890-91): "Notes sur le développement du langage", Revue de Linguistique et de Philologie Comparée 23, 330-343; 24, 10-42, 128-143, 242-257, 300-320.
- Ferguson, C. and C. Farwell (1975): "Words and sounds in early acquisition", Lg 51, 419-439.
- Ingram, D. (1974a): "Phonological rules in young children", Journal of Child Language 1, 49-64.
- Ingram, D. (1974b): "Fronting in child phonology", Journal of Child Language 1, 233-241.
- Ingram, D. (1978): "The acquisition of fricatives and affricates in normal and linguistically deviant children", in The Acquisition and Breakdown of Language, A. Caramazza and E. Zuriff (eds.), 63-85, Baltimore: Johns Hopkins University Press.
- Ingram, D. (in press): "Phonological patterns in the speech of young children", in Studies in Language Acquisition, P. Fletcher and M. Garman (eds.), Cambridge: Cambridge University Press.
- Ingram, D., L. Christensen, S. Veach and B. Webster (to appear): "The acquisition of word-initial fricatives and affricates in English by children between two and six", in Child Phonology: Data and Theory, J. Kavanaugh, G. Yeni-Konshian, and C. Ferguson (eds.).
- Klein, H. (1977): The Relationship between Perceptual Strategies and Productive Strategies in Learning the Phonology of Early Lexical Items, Doctoral dissertation, Columbia University.
- Priestley, T.M.S. (1977): "One idiosyncratic strategy in the acquisition of phonology", Journal of Child Language 4, 45-66.
- Roussey, C. (1899-1900): "Notes sur l'apprentissage de la parole chez un enfant", La Parole 1, 870-880; 2, 23-40.
- Vinson, J. (1915): "Observations sur le développement du langage chez l'enfant", Revue de Linguistique 49, 1-39.

THE ACQUISITION OF CHINESE PHONOLOGY IN RELATION TO JAKOBSON'S
LAWS OF IRREVERSIBLE SOLIDARITY

Heng-hsiung Jeng, National Taiwan University, Taipei,
Republic of China

I. Introduction

This paper attempts to find out whether the laws of irreversible solidarity as proposed by Jakobson (1968; 1971) also apply to the acquisition of Chinese phonology by two Chinese children.

Chinese here refers to the Mandarin Chinese as spoken in Taiwan, Republic of China, today. This variety of Mandarin Chinese is different from the standard Mandarin in that the former generally does not have the retroflex affricates /tʂ/, /tʂ^h/, and the retroflex fricative /ʂ/ that can be found in the latter. As far as tones and other segmental phonemes are concerned, they are essentially the same.

The subjects are my first son Jeng Wei, born on October 15, 1969, and my second son Jeng Hung, born on June 5, 1975. The data selected for this study are those of my first son recorded between the age of 2 months when babbling started and the age of 20 months when I left for the U.S. and stopped recording, and those of my second son recorded between the age of 15 months when he began to utter the first words and the age of 31 months when he had more or less mastered Chinese phonology. All these data were recorded by the author, mainly in phonetic transcription and occasionally with a tape recorder. My first son's data were mainly used for the discussion of the acquisition of tones, and the second son's data for the discussion of the acquisition of segmental phonemes.

My wife's native Chinese dialect is Hakka, and my native Chinese dialect is the South Min dialect spoken in Taipei. Most of the time we converse in the variety of Mandarin Chinese spoken in Taiwan as characterized above. Only when my wife's relatives or mine come to visit us is Hakka or South Min heard more often. So my sons generally live in the native-speaking environment of Mandarin Chinese, with only occasional exposure to Hakka and South Min.

II. Acquisition of Chinese Phonology

A. Tones

It has been observed by Jakobson (1968, 21-22) and Lenneberg (1964, 119) that babbling is not directly related to the acquisi-

tion of speech sounds. However, they did not touch upon the acquisition of tones in a tone language.

Chao (1951) noted that most Chinese children acquire tones quite early, except some tone sandhi phenomena. Li and Thompson (1976) and Li (1978) also observed that the acquisition of tones by a Chinese child precedes the acquisition of segmental phonemes. Weir (1966, 156) even pointed out that a Chinese baby at about six months, that is, during its babbling stage, had already much tonal variation over individual vowels, while the Russian and American babies at about the same age seldom showed such variation.

The written records of my first son's babbling show that he had tonal variation over individual vowels or syllables at a very early age: [ə/ə] (2 months); [e/ē] (3 months); [ɿ/ɿ̃] (3 months); [kuẽkuẽ] (3 months). And at 4 months, in response to my utterance [ãã], he produced a similar tonal variation over the same vowels. The above examples show that at the very early stage of babbling, he not only could produce vowels and syllables with different tones, but also link such production to the perception of tones.

This evidence supports Chao's (1951), Li and Thompson's (1976) and Li's (1978) observations that tones are acquired quite early by Chinese children. With such ability to perceive and produce tones transferred from the babbling stage, a Chinese child usually sets out to acquire his first Chinese words with practically correct tones. My first son, at 11 months, uttered his first word [papa] correctly with a falling tone followed by a neutral tone. At one year, he produced the word [pa^w] 'to mix milk powder with water' correctly with the falling tone even though the aspirated voiceless bilabial stop /p^h/ was incorrectly pronounced as its unaspirated counterpart. At 15 months, he could recognize the difference between [ɕiẽiẽ] 'shoes' and [ɕiẽiẽ] 'thanks', [xua] 'flower' and [xua] 'painting' because of their different tone patterns, even though he could not produce them yet. My second son at 16 months, about one month after the utterance of his first word, produced a minimal pair with tones as the distinctive elements: [pa^wpa^w] 'bread; food' and [pa^wpa^w] 'hold in arms'.

Mandarin Chinese has four tones and one neutral tone, which occurs only in an unstressed syllable. Besides the above mentioned high level tone [˥] (55), namely the first tone, in such a word

as [xua] 'flower', falling tone [˨˨˨] (51), namely the fourth tone, in such a word as [xua] 'painting', and neutral tone [˥˥] in the second syllable of the word [papa] 'father', there are the rising tone [˨˨˨] (35), namely the second tone, and the falling-rising tone [˨˨˨] (214), namely the third tone, which is realized as [˨˨˨] (35) when it occurs immediately before another third tone and normally realized as [˨˨˨] (21) elsewhere. Both my first and second sons acquired the second and third tones more or less simultaneously and without much difficulty: my first son had been able to produce the second-tone words [mai] 'buy', [niu] 'cow', and the third-tone words [tɕi] 'self' in [tɕi tɕi lai] 'by oneself', [tɕi] 'rise' in [tɕi lai] 'get up' by the age of 19.5 months; my second son uttered the second-tone words [tɕien] 'money' at 17.5 months, [nai] 'come' at 18.5 months, and the third-tone words [ta kai] 'open' at 16.5 months, [pa pe] 'urinate' at 17 months. However, they occasionally mispronounced some second-tone words as third-tone words and vice versa, and this supports the view of Li and Thompson (1976, 189) concerning such occasional confusion.

As for tone sandhi phenomena, the data of my second son, contrary to Chao's (1951) observations, show that he generally had no problem with them. And this also supports the view of Li and Thompson (1976, 189) that "tone sandhi rules are learned, with infrequent errors". For example, the third-tone word /uo/ 'I' before another third-tone word was correctly changed to the second tone in the expression [uo ie iau tsu tsy] 'I also want to go out' uttered at 21.5 months, while before a neutral-tone word, it was correctly realized as [uo] in the expression [uo tɕ] 'mine', uttered at 21.5 months. And the fourth-tone word /pu/ 'not' before another fourth-tone word was correctly changed to the second tone in the expression [pu tsai] 'absent' uttered at 24 months, but before a third-tone word, it remained unchanged in the expression [pu ɕi xuan] 'don't like' uttered at 22.5 months.

Therefore, Chinese tones, unlike segmental phonemes which have to evolve slowly step by step, are perhaps acquired by a Chinese child during babbling before the utterance of the first word and assigned immediately to the first words acquired.

But why are tones acquired before segmental phonemes? Perhaps the answer may be found in the lateralization of the human brain. Fromkin and Rodman (1974, 312) state that after lateralization,

the right brain is specialized in pattern-matching and the left brain in analytical thinking. Probably that is why such discrete linguistic elements as segmental phonemes can be acquired by the left brain after lateralization, which takes place around the age of one, and before lateralization, when both sides of the brain are still symmetrical, only suprasegmental patterns such as tones can be acquired.

B. Consonants and Vowels

My second son's acquisition of Chinese segmental phonemes may be divided into two stages: i. minimal phonological system; ii. fully developed phonological system, which is almost identical with an adult Mandarin speaker's phonology.

The minimal phonological system consists of four stops, /p/, which has two allophones [m] and [b] as free variants, /t/, /k/, and /ts/, which has an allophone [tʂ] occurring before /i/, and four vowels, /a/, /a^w/, /i/, and /e/. All these segmental phonemes were acquired within 44 days, between September 25 and November 8, 1976. And the words uttered within this period are as follows:

(9/25)¹ [pa_qpa_q] 'people'; (9/26) [ka^wka^w] 'older brother', [ie tɕi]~[te tɕi] 'eyes'; (10/4) [a tɕitɕi] 'dirty'; (10/12) [pa^wpa^w] 'bread; food'; (10/18) [pa^wpa^w] 'hold in arms'; (10/19) [tsatsa] 'dirty'; (10/26) [ta kai] 'open', [tata] 'candy', [te ta] 'fall down', [mapa]~[baba]~[papa] 'people'; (10/28) [piapia] 'don't want'; (10/29) [pa?pa] 'car'; (11/2) [βa] 'flower'; (11/7) [tia] 'drop'; (11/8) [pa pe] 'urinate'.

Beyond this stage of minimal phonological system, nasals, aspirated stops, fricatives except /f/, and the retroflex liquid /r/ emerged almost simultaneously although they became stable at different times. The lateral liquid /l/ appeared later than all these sounds, and /f/ was the last sound to appear. The following table shows when these consonants first emerged and when they became stable. The first number under each consonant indicates the age (in months) when it emerged, and the second number the age when

-
- (1) Hereafter, the Arabic numeral before a slash indicates the month and the Arabic numeral after it indicates the day of the month.
 - (2) This voiced bilabial fricative [β], which evolved into /x/ later on, does not fit into the minimal phonological system proposed here.

it became stable.

Table 1

Emergence and stabilization of further consonants in the fully developed phonological system

	p ^h	t ^h	k ^h	ts ^h	m	n	ŋ	r	f	s	x	l
Emer.	19	21.5	17	20	17	17	19	17	29	18	18	20.5
Stabi.	22.5	22.5	22	22	17	23	22.5	17	29	18	18	20.5

The vowels that emerged in the fully developed phonological system are /u/, /y/, /ɿ/, and /o/, which has the allophone [ɤ] when occurring after a nonlabial sound or occurring as a word-initial vowel. Once these vowels were acquired, they were very stable afterwards, except /y/, which at one time lapsed into [i^w] for the word "fish", whose proper pronunciation is [y]. The ages when these vowels appeared are given in the following table.

Table 2

Emergence of further vowels in the fully developed phonological system

	u	ɿ	y	o	ɤ
Emer.	17.5	20.5	19.5	18	20

The division of Jeng Hung's phonological development into the minimal phonological system and the fully developed phonological system may appear to be rather arbitrary. However, because of the simple distinctive features involved in the minimal phonological system and the complex distinctive features involved in the fully developed phonological system, the division is not without justification: in the minimal system, each of the consonants has only two distinctive features, that is, [+stop] and point of articulation, and each of the vowels is distinguished from the other vowels by two features, [±high] and [±low], except /a^w/, which has an additional feature of [+labialized]; in the fully developed system, after the age of 17 months, more consonants are distinguished by

more complex manner features such as [+aspirated], [+nasal], [±fricative], [±liquid], and [±retroflex] even though their points of articulation remain more or less the same as those of the stops in the minimal system, and vowels are further distinguished by [±back] and [±round].

III. Jakobson's Laws of Irreversible Solidarity

Jakobson (1968; 1971) set forth the laws of irreversible solidarity to account for the chronology of the acquisition of speech sounds by children, sound changes, and loss of speech sounds by aphasics. Now the acquisition of Chinese phonology by my sons Jeng Wei and Jeng Hung will be discussed in the light of his laws.

1) Jakobson did not touch upon the acquisition of tones in tone languages. According to Li and Thompson (1976) and Li (1978), the acquisition of tones precedes the acquisition of segmental phonemes. The discussion in II.A further points out that babbling has an important bearing on the acquisition of tones.

2) In the minimal phonological system of Jeng Hung, the vowels /i/, /e/, and /a/ form a vertical split as Jakobson predicted, but the labialized vowel /a^w/, which developed into the diphthong /au/ at 17.5 months, does not fit neatly into the pattern, and the consonants /p/, /t/, /k/, and /ts/ deviate from his laws of first and second consonantal split.

3) The early appearance of /k/ in Jeng Hung's minimal phonological system and /k^h/ and /x/ in his fully developed phonological system forms a counterexample to Jakobson's law that back consonants presuppose front consonants.

4) Jakobson's law that back rounded vowels presuppose their corresponding front unrounded vowels is supported by Jeng Hung's early acquisition of /i/ and /e/ and late acquisition of /u/ and /o/. So is his law that /y/ presupposes /i/ and /u/.

5) The almost simultaneous appearance of the aspirated stops, nasals, fricatives except /f/, and the retroflex liquid /r/ cannot be accounted for by Jakobson's laws. One tentative explanation proposed here is that these aspirated stops, nasals, and fricatives except /f/, being identical with their corresponding unaspirated stops in the minimal phonological system with respect to points of articulation, are developed simultaneously on the basis of adding to the existent unaspirated stops one more distinctive feature from the different manners of articulation such as [+aspirated],

[+nasal], and [+fricative]. The late acquisition of /f/, in the light of this explanation, may be due to the fact that its point of articulation is different from any of the unaspirated stops in the minimal phonological system, hence the substitution of /f/ by the voiceless bilabial fricative [ɸ] in the words [i₁ ɸu₁] 'clothes' and [ɸei₁ tɕi₁] 'aeroplane'. As for the simultaneous acquisition of /r/ with aspirated stops, nasals, and fricatives except /f/, one possible explanation is that the additional distinctive feature of [+retroflex] is combined with the negative values of these manners of articulation as a clear-cut opposition.

6) Jakobson (1968) pointed out that the second liquid is one of the last sounds acquired by the child. The late acquisition of /l/ by Jeng Hung at 20.5 months, only before /f/, the last sound acquired, supports his view.

References

- Chao, Y.R. (1951): "The Cantian idiolect", Semitic and Oriental studies presented to William Popper, University of Calif. Publications in Semitic Philology II, 27-44.
- Fromkin, V.A. and R. Rodman (1974): An introduction to language, New York: Holt, Rinehart and Winston.
- Jakobson, R. (1968): Child language, aphasia and phonological universals, The Hague: Mouton.
- Jakobson, R. (1971): Studies on child language and aphasia, The Hague: Mouton.
- Lenneberg, E.H. (1964): "Speech as a motor skill with special reference to non-aphasic disorders", in The acquisition of language, U. Bellugi and R. Brown (eds.), 115-127.
- Li, C.N. and S.A. Thompson (1976): "The acquisition of tone in Mandarin-speaking children", Journal of Child Language 4, 185-199.
- Li, P.J.K. (1978): "Child language acquisition of Mandarin phonology", in Studies and essays in commemoration of the golden jubilee of the Academia sinica, 615-632.
- Weir, R.H. (1966): "Some questions on the child's learning of phonology", in The genesis of language, F. Smith and G.A. Miller (eds.), 153-172, Cambridge: The MIT Press.

PREDISPOSITIONS FOR THE PERCEPTION OF SPEECH BY HUMAN INFANTS
Patricia K. Kuhl, Department of Speech and Hearing Sciences,
 Child Development and Mental Retardation Center, University of
 Washington, Seattle, WA. 98195.

The development of speech production and perception in the human infant shares certain themes with the acquisition of communicative repertoires in animal species. Among those themes is the notion that infants of a species demonstrate predispositions for the perception of communicatively relevant acoustic signals. While the animal literature provides examples in which innate predispositions are in evidence, a growing body of literature on the complex role of "normal" experience, and the effects of selective auditory exposure, in maintaining, facilitating, and inducing such behavior is accruing, leading to the hypothesis that infants are predisposed toward fairly simple acoustic features and develop the perception of "configurational" models only with experience. Two approaches to examining the role of experience in the perception of speech by human infants are discussed.

Converging Themes in Developmental Neurobiology

At the end of the first decade of research on the perception of speech by young infants, the list of published experiments is long and the speech features that have been examined is extensive (see Kuhl, In Press, for review). The common theme running through this work is the examination of potential auditory perceptual predispositions that human infants bring to the task of learning language - predispositions that would direct the infant toward the acoustic features that are particularly relevant to the perception of speech, such as those acoustic features which signal the segmental and nonsegmental elements of the language.

The notion that members of a species may be perceptually predisposed to attend to, resolve more precisely, respond to, or to otherwise treat differently, visual and auditory signals that are relevant to their survival is an old theme in the literature on communicative behavior in animals and humans (Lorenz, 1965). Many attribute stimulus prepotencies to species-specific neural mechanisms that have evolved specially for that purpose and perceptual predispositions that are innately determined. The evidence for such mechanisms is both behavioral and physiological

and largely stems from work on animal communication (see Schneich, 1977, for a review of neurophysiological data and Gottlieb, 1976a, for a review of behavioral data).

The discovery in behavioral and physiological experiments that communicatively relevant stimuli enjoy special status for the adult perceiver naturally raises questions about the development of these behaviors in infants of the species. While the early theorists (Lorenz, 1965; Tinbergen, 1951) stressed the "instinctiveness" of certain behaviors and underplayed the role of experience, "learning" in the classical sense, or maturation in the development of complex behavior, more recent theorists (Gottlieb, 1976a) have stressed the complex role that experience plays and the variety of different ways experience affects the organism (Gottlieb, 1976b).

Recent physiological evidence suggests that sensory input during early development has an effect on central neural mechanisms, particularly in the visual system; the responsiveness of units in the visual cortex of adults is biased by distorting or denying early "normal" visual experience, or by selective visual exposure. This physiological "plasticity" in the visual system can be species-specific and evidence for "critical periods" exists (see Daniels and Pettigrew, 1976, for review).

The effects of selective auditory exposure are less well known. Silverman and Clopton (1977) and Clopton and Silverman (1977) noted substantial losses in binaural interaction at the inferior colliculus in rat after early monaural deprivation. Clopton and Silverman (1978) demonstrated changes in the latency and duration of neural responses to clicks at the level of the inferior colliculus in rat after early auditory deprivation. Clopton and Winfield (1976) further demonstrated using the rat that exposure during the first four months of life to patterned sound (upward tone sweeps, downward tone sweeps, or noise bursts) increases the response of units in the inferior colliculus to that pattern relative to a similar but inexperienced pattern. No effects of selective exposure were found in an adult population of rats.

Perhaps the best examples from the animal literature on the interactions between innate predispositions and experience are

to be found in the growing literature on song learning in the Passerine bird (Marler, 1973). Certain songbirds must hear their songs in order to learn them but there are interesting constraints on learning; the exposure must be to the conspecific song and it must occur during a "critical period." Marler hypothesizes that song vocalization is developed by reference to an "auditory template," a mechanism that is specific enough to detect some of the critical features of the conspecific song and thus direct the bird's attention in its direction, but one which requires exposure to the song to "fill in" the details of its acoustic structure. As Marler (1973) describes, their learning is not left purely to chance, it ". . . takes place within a set of constraints which seem designed to ensure that the learning bird's attention shall be focused on a set of sounds that is biologically relevant. . ." (p.80). To make the songbird parallel even more striking, Nottebohm *et al.* (1976) have demonstrated functional hemispheric asymmetry for the production of song in these birds. Using ablation techniques, they have demonstrated that the left hemisphere controls song production in the Canary, but if ablation of the left motor area occurs before the bird has passed the critical period for vocal learning, the bird's song develops normally using the subordinate right motor area.

Marler interprets these data as indicating that the innate direction that the infant comes into the world with is simply that - a direction or guideline pointing the infant in the appropriate direction, rather than a complete "schema" of the song. He believes that the predispositions are toward rather simple stimulus features and only with continued exposure to the configuration that is being detected does the infant develop a "schema" of the complex stimulus array.

Predispositions for the Perception of Speech by Human Infants

There are two ways in which the role of experience is currently being examined for the perception of speech by human infants. One approach is to chart the course and examine the nature of perceptual changes that occur as a result of exposure to a particular language. Another approach is to examine the infant's recognition of abstract auditory-phonetic categories rather than simple stimulus features, expecting that the former

may reveal developmental trends.

How does linguistic exposure modify the way in which infants perceive speech sounds? While not well understood, the perceptual effects of exposure to one's native language have been documented in adult listeners (Miyawaki *et al.*, 1975; Abramson and Lisker, 1970). Taken together with the existing data on the perception of speech by infants, these data have led to the hypothesis that infants discriminate all of the simple phonetic contrasts at birth regardless of their linguistic environments, but that due to the lack of exposure to certain phonetic units during development the infant somehow loses the ability to distinguish them from contrasting phonetic units.

Attempting to chart developmental changes in an infant's perception that can be attributed to linguistic exposure has received some attention, but we are still without a simple answer to the question. The evidence is fairly convincing that infants being reared in non-English-speaking environments are capable of discriminating at least one phonetic contrast (voiceless-unaspirated /pa/ from voiceless-aspirated /p^ha/) that is phonemic in English but not in the infant's native language. Streeter (1976) using the sucking-habituation technique, demonstrated that two-month-old African Kikuyu infants discriminated the English contrast in addition to discriminating a voicing contrast that is phonemic in the Kikuyu language but not in English (prevoiced /ba/ from voiceless-unaspirated /pa/). Lasky, Syrdal-Lasky and Klein (1975) demonstrated similar results for Spanish infants of the same age using a heart-rate technique.

On the other hand, the case for discrimination of the prevoiced /ba/ from the voiceless-unaspirated /pa/ by American infants is not quite as clear. Recent studies (Eilers, Wilson and Moore, 1977; Eimas, 1974) have failed to provide evidence that American infants discriminate pairs of stimuli that are as close on the continuum as those discriminated by the Spanish and Kikuyu infants. However, there are a number of problems with these cross-language comparisons. First, the stimuli are synthesized to manipulate an acoustic cue that is acoustically fragile and is likely to be subject to variation due to the differences in acoustic calibration across laboratories. A more recent set of

studies claims to be immune to this criticism. Using the head-turn technique, Eilers, Gavin and Wilson (In Press) tested six-month-old American and Spanish infants in the same laboratory, but in two different studies, and demonstrated that while both groups discriminated the English contrast, only the Spanish infants discriminated the Spanish contrast.

Do infants recognize the configurational properties of phonetic categories? Only recently have researchers attempted to find out whether infants are capable of recognizing the similarity among sounds that have the same phonetic label when the sounds occur in different phonetic contexts, when they occur in different positions in a syllable, or when they are spoken by different talkers.

A conditioned head-turn response for visual reinforcement has been successfully used with six-month-old infants to test the recognition of phonetic categories (Kuhl, 1978). In these tasks, infants are trained to make a head-turn response when one speech token is changed to another speech token (like from /a/ to /i/). During training, vowels produced by a male talker (computer-synthesized) are used; subsequently, infants are tested with computer-synthesized vowels produced by female and child talkers. The ease with which the infant generalizes to new exemplars from the category indicates the degree to which the infant perceives the similarity among the tokens from a given category.

Results to date in these category-formation tasks strongly suggest that vowel categories are readily perceived by the infant listeners. Tasks requiring the infant to recognize a change from the vowel category /a/ to the vowel category /i/ and tasks requiring the infant to recognize a change from the vowel category /a/ to the vowel category /ɔ/ result in near perfect transfer of learning to the new tokens from the categories (Kuhl, 1978). We have also completed studies on the categorization of fricative consonants, such as /f/ vs. /θ/, and /s/ vs. /ʃ/ (Holmberg, Morgan and Kuhl, 1977). In general, our results suggest that the /a-i/ contrast is the easiest in this category-formation task, that the /f-θ/ contrast is the most difficult one, and that the /a-ɔ/ and the /s-ʃ/ contrasts are of intermediate difficulty.

These category-formation experiments (discussed in detail in Kuhl, 1978) have two advantages. First, one can test the infant's recognition of abstract configurational properties of speech-sound categories, and second, one can test how readily or efficiently the infant forms categories based on dimensions that are not phonetically relevant, at least in English, such as pitch contour or stress. These techniques may demonstrate that all infants recognize categories based on certain "focal" auditory dimensions, but that their tendencies to attend to particular acoustic dimensions is modified by exposure to a particular language.

Systematic experiments examining the perception of abstract perceptual categories, rather than simple discriminations, in at least two different populations in which the target acoustic features are chosen such that they are phonemically relevant to one population and not to the other are necessary before the contributions of innate predispositions and experience will be understood in the development of speech perception.

References

- Abramson, A. and Lisker, L. (1970): "Discrimination along the voicing continuum: Cross-language tests," in Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967, Academic, 569-573.
- Clopton, B.M. and Silverman, M.S. (1977): "Plasticity of binaural interaction. II. Critical period and changes in midline response," J. Neurophysiol. 40, 1275-1280.
- Clopton, B.M. and Silverman, M.S. (1978): "Changes in latency and duration of neural responding following developmental auditory deprivation," Exp. Brain Res. 32, 39-47.
- Clopton, B.M. and Winfield, J.A. (1976): "Effect of early exposure to patterned sound on unit activity in rat inferior colliculus," J. Neurophysiol. 39, 1081-1089.
- Daniels, J.D. and Pettigrew, J.D. (1976): "Development of neuronal responses in the visual system of cats," in Neural and Behavioral Specificity, Vol. III, G. Gottlieb (Ed.), 196-232, New York: Academic Press.
- Eilers, R.E., Gavin, W.J. and Wilson, W.R. (In Press): "Linguistic experience and phonemic perception in infancy," Child Develop.
- Eilers, R.E., Wilson, W.R. and Moore, J.M. (1977): "Speech discrimination in the language-innocent and the language-wise: A study in the perception of voice-onset time," J. Acoust. Soc. Am. Suppl. 1, 61, S38(A).

- Eimas, P.D. (1974): "Linguistic processing of speech by young infants," in Language Perspective-Acquisition, Retardation and Intervention, R.L. Schiefelbusch and L.L. Lloyd (Eds.), 55-74, Baltimore: University Park Press.
- Gottlieb, G. (1976a): "Early development of species-specific auditory perception in birds," in Neural and Behavioral Specificity, Vol. III, G. Gottlieb (Ed.), 237-281, New York: Academic Press.
- Gottlieb, G. (1976b): "The roles of experience in the development of behavior and the nervous system," in Neural and Behavioral Specificity, Vol. III, G. Gottlieb (Ed.), 25-54, New York: Academic Press.
- Holmberg, T.L., Morgan, K.A. and Kuhl, P.K. (1977): "Speech perception in early infancy: Discrimination of fricative consonants," J. Acoust. Soc. Am., Suppl. 1, 62, S99(A).
- Kuhl, P.K. (1978): "Perceptual constancy for speech-sound categories in early infancy," Chapter presented at the NIH Conference on Child Phonology, May, 1978.
- Kuhl, P.K. (In Press): "The perception of speech in early infancy," in Speech and Language: Research and Theory, N.J. Lass (Ed.), New York: Academic Press.
- Lasky, R.E., Syrdal-Lasky, A. and Klein, R.E. (1975): "VOT discrimination by four- to six-and-a-half-month-old infants from Spanish environments," J. Exp. Child Psych. 20, 215-225.
- Lorenz, K. (1965): Evolution and Modification of Behavior, University of Chicago Press, Chicago.
- Marler, P. (1973): The Clarence M. Hicks Memorial Lectures for 1970, University of Toronto Press, Toronto, 69-85.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins J., and Fujimura, O. (1975): "An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English," Percept. Psychophys. 18, 331-340.
- Nottebohm, F., Stokes, T.M. and Leonard C.M. (1976): "Central control of song in the Canary," J. of Comp. Neurol. 165, 457-486.
- Schneich, H. (1977): "Central processing of complex sounds and feature analysis," in Recognition of Complex Acoustic Signals, T.H. Bullock (Ed.), 161-182, Berlin: Abakon Verlagsgesellschaft.
- Silverman, M.S. and Clopton, B. (1977): "Plasticity of binaural interaction. I. Effect of early auditory deprivation," J. Neurophysiol. 40, 1266-1274.
- Streeter, L.A. (1976): "Language perception of 2-month-old infants shows effects of both innate mechanisms and experience," Nature 259, 39-41.
- Tinbergen, N. (1951): The Study of Instinct, Clarendon Press, Oxford.

TRANSITION AND VARIATION IN CHILD PHONOLOGY: MODELING A
DEVELOPING SYSTEM

Lise Menn, Aphasia Research Center, Boston University
School of Medicine, Boston, Massachusetts, U.S.A.

Child phonology is different from general phonology in several important areas. When we try to characterize those differences we find in many cases that a set of phenomena which play a central role in the one field play a marginal role in the other. It is quite reasonable a priori that this should be the case when we consider the topic of variation in child phonology versus the topic of variation in adult phonology: the very notion of acquisition implies long-term change in performance, whereas we assume that in the adult, the phonology is sufficiently stable for any change to be relegated to the limbo of marginal phenomena.

In this paper, I will briefly review certain types of variation which are prominent in child phonology, and consider how one might incorporate these types of variation in a theoretical model. We will not take up those types of variation that are prominent in both child phonology and adult phonology, such as registral, sociolinguistic, allomorphic, and allophonic variation, although a complete model must deal with those as well; we will keep to the more restricted topic of those types of variation that seem to be intimately associated with the process of the acquisition of phonology.

These will include, as mentioned, long-term changes in rules and pronunciations. These are orderly, one-way transitions in language behavior: the child learns to hit a particular phonetic target, or learns to render a particular sequence of sounds in accord with the adult model word instead of producing it in some scrambled order.

Acquisition studies show that there are also several types of short-term variation among renditions of a given word. Two of these can be considered as being the microstructure of long-term variation: transitional variation and local scatter in the production of a particular phone in a phonologically defined context.

Transitional variation refers to the vacillation between

well-defined pronunciations of a word that frequently occurs during the period when an old rule is being superseded by a new rule. Such bimodal variation in renditions of a word is usually taken as evidence that two rules are in conflict. Sometimes the changeover from old to new rules has an intermediate period showing transition variation, and sometimes no such period is observed.

Local scatter is a unimodal variability in the production of a particular phone. This simply looks like the result of poor articulatory control compared to the adult norm: the child's shots at a target more often fall wide of the mark. (There must also be a second-order long-term variation associated with local scatter, since we expect to see a reduction in local scatter as the child matures.)

Presently I can enumerate five other kinds of short-term variation. One of these is called backgrounding (Ferguson & Farwell 1975). As they say, one portion of a word may be "deleted or drastically reduced while the child is 'working on' another part of the word." They cite from their data one child's production of 'milk' as [bʌʔ] and [ʌk̄] in the same session. I think we now have enough evidence from selective avoidance (Ferguson & Farwell 1975) to assert that children can and sometimes do monitor the quality of their own output; therefore, the most reasonable explanation of backgrounding as Ferguson & Farwell describe it is to assume that it takes place under conditions of high self-monitoring of the phonetics of the output or the input.

A second type of variation which also seems to involve self-monitoring is the well-documented imitation effect: a word may be pronounced very differently when it is an imitation than when it is produced without the adult model ringing in the child's ears. Frequent anecdotes report one sub-type of model-induced variation: a child will be reported to have said a word 'perfectly' or nearly so on the very first attempt, and then to have reduced it drastically in later renditions. One would expect to find parallels to backgrounding and imitation-effect variability in adult speech when one is attending to the sound of the word as well as its meaning, while speaking.

(It is also well-known that children can spectacularly fail to be aware of the sound of their output, and imitation may fail

to induce any variation at all; he or she may insist vehemently that what she/he said is the same as what the modeling adult has said. It is of course difficult to know whether the child is referring to pronunciation or to content in such assertions; metalinguistic conversations with two-year-olds tend to be unsatisfactory (Brown & Bellugi, 1964; in Brown, 1970, p. 79.)

The third type of unexpected variation is again a bimodal variation brought about by rule conflict, but this time it is not a passing unstable phase marking the cusp-point of change. Instead, it seems to reflect the co-existence of competing rules which may arise and decay at about the same time (Menn, 1973). We will refer to this as rule-coexistence variation when it is necessary to distinguish this type of rule-conflict variation from transition variation.

A fourth interesting kind of variation, which we will call floundering, can be described as wide fluctuation in the production of a particular model phone or string of phones under phonologically stable conditions. An example is Daniel Menn's 'peach' attempts, [itš] [dits] [pipš] [gik] [nitš] etc. (Menn 1973). This kind of variation I have interpreted as being what happens when a child has no well-formed rule for dealing with a particular string of phones, that is, where the model word does not meet the structural description of any of the child's rules, and where the outputs look like what would happen if one or several features of the model word were changed so that it could be an input to the child's rules. Conceptually, floundering is quite distinct from backgrounding; floundering is the result of trying to use rules that don't quite apply, while backgrounding occurs when the child's output is produced with less reliance on practiced rules and more attention to pronunciation as a task. The parallel distinction can be made in adult second-language learning. Suppose we have an American trying to pronounce a hypothetical word /ndaŋa/, containing the unEnglish cluster /#nd/ and the morphologically controlled medial /ŋ/. Suppose our speaker is able to get each of these difficult items correct when thinking about it, but that s/he otherwise reverts to initial /#end/ or /#d/ and to medial /ŋg/. The variation between /ŋ/ and /ŋg/ (and also between /#nd/ and either of the two wrong pronunciations) is controlled by the amount of attention that it gets: this is

backgrounding. On the other hand, the variation between /#end/ and /#d/ for /#nd/ is floundering: it is a random choice among sounds which have a close resemblance to the difficult target.

Finally, some young children show lexically controlled variation. Here, certain words show great variation in the production of some or all their sounds while other words that have similar adult models show much less variation. Jacob (Menn 1976) had a much greater variability for the /æwn/ sequence in 'down' than for the same target in 'around'. This also has parallel at the margins of adult phonology: consider for example the great variety of sounds permissible (as expressive variants) for the 'phoneme' /o/ in the word 'no'. This variety is not found in renditions of the same phoneme in the word 'know'.

We have named seven types of variation of special interest to child phonology. Now, by a 'model' of a phonological system, I mean a flow chart which specifies roughly what information is stored, what is used in real time, and how the different pieces are brought together to specify the articulatory instructions needed to produce a word. How can these seven types of variation be represented in such a model?

The most important capability to be added to extant models actually is, I think, one that has not been explicitly mentioned so far, since it manifests itself indirectly. Child phonology models almost all represent the steady state: the rule or word is established. These models need new apparatus to simulate what happens when a new word is being tried or a new rule is being formed, for practiced behavior is very different from novel behavior. This familiar-novel distinction seems to be related to the distinction that we have already invoked between monitored and automatic behavior, but they are not the same. To deal with both novelty and attention, models will have to allow more than one route from adult word to child word. We could say that one route, the one used most frequently, would represent automatic, over-learned behavior, and other routes would correspond to the special cases when at least part of a word is not being produced under automatic control. We can make this more explicit by considering an available child-phonology model.

Suppose we use a two-lexicon model similar to the one in Kiparsky and Menn (1977), concerning ourselves with the part of

it that would run: adult form + (perceptual strategies) + phonetic representations perceived by child = input lexicon + (reduction rules) + encoded articulatory representations = output lexicon + (motor routines) + child's output form. The input lexical entry represents the child's encoding of his/her percept of the adult word, the output lexical entry represents an encoding of articulatory instruction, and the reduction rules relate the two lexicons.¹ We can modify such a model to allow for non-automatic speech production by adding routes from the input lexicon (percept of model word) to the output side (pronunciation) that bypass the output lexicon and some of the rules that lead into and out of it. This would represent an attempt to give a spontaneous rendition of a known word without most of the automatic apparatus, and might represent what goes on during word-practice. To represent imitation, we would also add routes from some point(s) among the perceptual processing routines that would bypass both lexicons and feed into some points among the articulatory routines.

The variation in the points of beginning and ending of these bypasses would reflect the degree to which established perceptual and articulatory routines were employed in the utterance. (Presumably, the more that one monitors, the more habits of perception and production can be overcome.)

It seems, then, that some aspects of transition (rule change), backgrounding, and imitation-effect variation can be modeled by the addition of these new processing 'routes' to a K & M-type model. It turns out only a few more entities are required to adapt this model or its descendants to represent the other four types of variation that we have discussed.

Coexistence variation can be simulated by letting both of the competing reduction rules operate on each applicable input lexical item, thus generating two forms in the output lexicon corresponding to each of those input forms. Either of those forms could be translated into output any time the child said the word. If the probabilities that the two forms both occur are not equal, some notion of the 'strength' of a lexical entry must also be added, so that one could say that the stronger entry is the

(1) The recent revision of the K & M model presented in Menn 1977 would allow a clearer formulation of some of the following discussion, but occasions no major differences.

one produced more frequently.

Transition variation would also be represented by having two output lexical entries, one generated by the older rule and one by the new rule. As we have implied, we can model the loss of a rule by removing it from the set of production rules. This will 'disconnect' some output lexical entries from their input lexical entries. (Transition variation would thus not be rule competition, as we stated above, so much as competition between two output lexical entries.) Since new rules normally spread to older words, we might hypothesize that the 'disconnected' output lexical entries lose strength and fade away. However, we know that some lexical entries which clearly do not have live support, such as phonological idioms and fossils (words which inexplicably resist rule changes), do not fade in the usual way but remain vigorous for long periods. If the 'fading' notion is used, we require special apparatus to handle phonological idioms and fossils. Several have been proposed (see Macken 1978) but we cannot pursue that topic here.

Local scatter does not involve lexical entries at all, but has to do with the lowest output processing levels: we shall assume that it occurs when articulatory instructions for a phone are executed with more tolerance than they would be by an adult.

Lexically controlled variation, on the other hand, requires, obviously, a special entry in the output lexicon just as phonological idioms do, and in addition this entry must specify special articulatory instructions rather than the general output routines or in addition to them.

The remaining form of variation that we have discussed is floundering. The basic situation in floundering seems to be a rule-input that is ill-formed. The proper analysis of a given case, however, may depend on the whole rule-structure, because there are several ways that this could happen in the present type of model. There are two relevant loci: the input lexical entry could fail to meet the structural description of necessary reduction rules, or the output lexical entry could fail to be of the proper form for the articulatory instructions to handle. In addition either case of ill-formedness might be better modeled by overspecification, underspecification, or some other type of malformation. Further elaboration of the psychological interpre-

tation of this or similar models of child phonology will be required in order to make a principled choice among these alternatives.

To conclude: certain types of variation are intimately and essentially involved with learning to pronounce. As we build richer models of child phonology, we can incorporate them without undue difficulty. Regardless of how easily we can draw new lines and little boxes, however, one problem about transition and variation remains very difficult. How does a new linguistic behavior cease to be effortful and become automatic?

References

- Brown, R., and U. Bellugi (1964): "Three processes in the child's acquisition of syntax", in Psycholinguistics, R. Brown, Glencoe, Ill: The Free Press.
- Ferguson, C.A. and C.B. Farwell (1975): "Words and sounds in early language acquisition", Lg 51:419-439.
- Kiparsky, P. and L. Menn (1977): "On the acquisition of phonology", in Language learning and thought, J. Macnamara (ed.), New York: Academic Press.
- Macken, M.A. (1978): "The child's lexical representation: the puzzle-puddle-pickle evidence" (ms.), Stanford University Linguistics Department.
- Menn, L. (1971): "Phonotactic rules in beginning speech", Lingua 26:225-251.
- _____ (1973): "On the origin and growth of phonological and syntactic rules", Papers from the Ninth Regional Meeting of the Chicago Linguistic Society, 378-385.
- _____ (1976): "Pattern, control, and contrast in beginning speech: a case study in the development of word form and word function", U. Illinois doctoral dissert., to be circulated by the Indiana University Linguistics Club.
- _____ (1977 and to appear): "Phonological units in beginning speech", in Syllables and segments, A. Bell and J. Hooper (eds.), Amsterdam: North Holland Publishing Co.

SPEECH SOUND CATEGORIZATION BY CHILDREN

Paula Menyuk, Applied Psycholinguistics, Boston University,
Boston, Mass., USA

Clearly acquisition of the structural properties of language takes place via a process of segmenting and then categorizing the segments of the language heard into units which can be used to comprehend and generate unique utterances. Equally clearly, the first aspect of language that is used by the hearing adult and by the infant to segment and categorize utterances is the acoustic speech signal. Unlike the adult, however, who has already determined what the "appropriate units" are and can by-pass much of the surface structure of the utterance, the infant must rely heavily on the signal to come to conclusions about appropriate segmentations. She must also rely heavily on contextual cues to relate the segments of the signal to objects and events in the environment. Despite the fact that common sense tells us that the above must be the case, we are, at the present time, still unclear about what these segments are and what the bases for categorization of segments are either initially or over time as the child matures. Indeed, controversies still exist in the literature on this issue for the adult as well as for the child. In this paper varying hypotheses concerning the nature of and bases for speech sound categorization by the child and its role in language acquisition will be examined in light of theories on adult language processing and the data on the speech processing behavior of the infant and young child.

Theoretical Descriptions of Processing

The segmentation of continuous speech has been described by linguists as being hierarchical and nested. That is, a message can be characterized as a type of speech act. The message can contain a sentence or sentences and these contain phrases which are made up of morphemes. Morphemes are composed of syllables which are made up of speech sound segments each of which represents a bundle of features. If this were a psychologically real description of language processing as well as of elements of language then the listener would determine categories of segments in a sequence with each lower step of the sequence dependent on the immediately higher step since higher steps indicate units of analysis. Speech sound segment categorization would take place at the end of the sequence

and require resolution of the bundle of features. Varying descriptions based on this hierarchical model have been labelled "analysis by synthesis" (Cooper, 1972). It might be logically argued from this model that since speech sound categorization or identification is comparatively late in the sequence of on-line processing then it must also be late in the sequence of acquisition of the structural properties of the language.

The above model of language processing has been deemed inadequate in accounting for either real-time processing of speech by the adult or for the observed sequence of acquisition of structural properties by the child. Since earliest utterances are sequences of speech sounds marked prosodically and the infant does not appear to understand anything more about utterances than their affective intent, it cannot account for behavior during the early babbling period. The model does not account for subsequent language behavior since even then the child does not evidence any knowledge of any of the postulated higher categories (i.e. sentence, phrase and morpheme). An alternative description of both processing and the sequence of acquisition is a bottom-up model or "synthesis by analysis". With this model speech sounds are differentiated and categorized, then grouped into higher level categories in a sequential manner during speech processing. In acquisition, speech sounds are differentiated and categorized by a process of imitation and sound approximation which is rewarded. These sounds are then composed into words by the same process and by associating phonological sequences with objects and events. Larger units of an utterance, phrases and sentences, are composed by putting together smaller units via a chaining process (Staats, 1971). This description suggests that the earliest analysis in processing and the earliest structural acquisition are segmental speech categories although the nature of these categories is not defined in the model.

Not only is there a substantial amount of evidence to indicate that this model does not adequately describe adult language processing (Fodor et al., 1974), but, also, it is difficult to see how the analysis of speech sounds one by one in a sequence can lead to decisions about where crucial boundaries lie and, thus, to a determination of meaning. For these same reasons a synthesis by analysis model seems inadequate in accounting for language acquisition. Although there is evidence that in early care-giver-child

communicative interaction, segments and boundaries are made much more salient than they are in adult-adult communication (Newport, 1976), there is no evidence that the child, in the process of acquisition, adds sequentially to segments by chaining bits together or that the child merely imitates input structures. On the contrary, the child appears to be only able to attend to and generate certain aspects of utterances at certain periods of development regardless of input structure and these aspects are not sequential bits of adult utterances.

Still another description suggests that perception and generation of connected speech is a parallel process; i.e. one involving all components of the language simultaneously. In the process chunks of the message, probably phrases, are subjected to analysis and rough estimates are made of the phonological composition of the morphemes in the phrase to corroborate hypotheses about the meaning of the phrase and then other phrases if more than one is contained in the utterance. An exact representation of the phrase can be kept in mind until analysis is completed so that needed corrections on this estimate can be made (Garrod and Trabasso, 1973). The child, during the process of acquisition, would analyze the data in the same fashion. The distinctions between the child and the adult are in the amount of information chunked for analysis, the much heavier reliance on the part of the child on contextual cues for analysis and the process of chunking itself since segmentation strategies would change as more structural knowledge of the language is acquired (Menyuk, 1977, Chap. 5). For example, an early chunking strategy might be to ignore everything in the signal except those sequences that signal main relations of actor and action or action and object. Components of the relation would be grossly analyzed for lexical look-up. However, analysis of the phonological segments per se would not be needed for comprehension. Since the parallel processing requires analysis of segments only when correction of rough estimates is required it might, again, be logically argued that speech sound segment categorizations would be later acquisitions than morpheme categorizations.

The above are theoretical descriptions of adult language processing and theoretical descriptions of the process of language acquisition. In conjunction with these are descriptions which are concerned with phonological acquisition only. This acquisition has

been described as a process of first discriminating between the speech sounds of the language, then categorizing these distinctions in terms of articulatory gestures. These discriminations are based on distinctive feature differences between speech segments. Early distinctions are determined by feature detectors that are pre-programmed in the auditory system of the human infant (Eimas, 1974). These might be termed primary features. Finer distinctions are then made both in perception and production and are probably affected by particular language experience. However, given the universality of the speech processing abilities of normal infants, both perceptually and productively, there is, to some extent, universality in the sequence in which distinctions are made. This universality is confounded by the particular data the child is confronted with; i.e. the language of the child's community and even family. Thus, the universal order is modified by the perceptual and productive problems a particular language poses for the child and by the interactive styles, lexical selections, etc. of a particular family. Individual differences become more marked when standard lexical items in a particular language begin to be used.

Data on Early Speech Processing

On the face of it there appears to be a logical gap between theories of language processing and of language acquisition and theories of the development of the phonological system. The latter suggest very fine analysis of the signal on the segmental level in terms of distinctive features, whereas the former suggest rather gross analyses dependent on higher level categories. There are also large differences between the findings of studies carried out at different periods of early speech processing behavior. One of the primary reasons for these gaps between theories of language and speech acquisition and between the findings of studies of speech processing and the conclusions drawn from them may be not maintaining a clear distinction between what the infant and child can do and what they ordinarily do; i.e. a capacity versus performance distinction. The data collected thus far on early speech processing indicate that the very young infant (1 to 4 months) as well as the very young child (under two years) can discriminate between speech sound segments that vary in terms of a single distinctive feature. There also appears to be a hierarchy in the features that can be distinguished both perceptually and productively. Thus, during

the cooing and babbling periods some features appear to be more perceptually salient than others and this also appears to be the case when the task is distinction of minimal pair nonsense syllables. Similarly, segments containing certain features are realized before segments containing other features in babbled utterances and then in morpheme production. However, there is not an exact correlation between the order of perceptual and productive distinctions made, and individual differences in the exact sequence of features and segments distinguished can be observed.

What seems to be suggested by these data is that distinctions can be made on the basis of distinctive features by the infant and young child if the question is put to them in a way in which these distinctions are made clear; i.e. in a small enough context that is non-distracting such as nonsense-syllables. Also, the response required in the task must be part of the children's behavioral repertoire. For example, they must have sufficient memory to recall the stimuli presented. Finally, there are some features that can be distinguished before others. However, discrimination between features does not imply that categorization of segments has taken place in terms of bundles of features nor does the capacity to discriminate between features imply that this is what children do when they listen to speech and attempt to match articulatory outputs to stored representations. Indeed, all the data indicate that during the babbling and early lexical acquisition periods distinctive feature differences are not actively employed in determining meaning of utterances or in generating utterances.

During the babbling period perceptual processing of continuous speech seems to be primarily based on the supra-segmental aspects of the speech signal and contextual cues. Some time toward the end of this period recognition of a small set of lexical items is observed and still later production of word approximations begins. The lexicon of the child at this time is quite small. It is entirely reasonable to suppose that both lexical recognition and generation are based on syllabic representations of morphemes. In other words, speech processing is taking place on the basis of the morpheme and this may be the minimal unit for categorization of speech information. The meaning of a phonological sequence, its gestalt phonological representation as a syllable or reduplicated syllables, supra-segmental features of intonation and contextual cues appear

to be all that is needed or used to comprehend or generate utterances during this time (Menyuk and Menn, in press).

Again, this is what children appear to do in on-line processing of speech during these early periods of development, although, at this time and long before, they are capable of discriminating between speech sounds on the basis of feature distinctions. As the lexicon grows and as structural knowledge increases constraints on memory probably make segmental differentiation and categorization necessary. When this occurs an available competence is actively employed. However, segmental differentiation and categorization may be needed only rarely to comprehend continuous speech. Thus, although the ability may be increasingly used at later periods of development it still may be used infrequently. Research shows that even 3 and 4 year-old children first use morpheme information to differentiate between phonological sequences and only use segmental information with some exertion when morpheme information is unavailable; i.e., with nonsense syllables or unknown words. At present, little is known about when reference to segmental information is used without marked exertion. Such ability is, of course, required in learning to read alphabetic text. One would assume that this ability develops gradually and that there would be individual differences or group variations due to language experience in the ages at which this ability manifests itself (Savin, 1972).

Conclusions

The theoretical description of the processing of language which appears to most adequately describe the sequence of acquisition of the structural properties of the language and to best fit the data on infants and young children's speech processing is one of parallel analysis of chunks of continuous speech. Initially the chunks the child can process are short in duration, linear in arrangement and involve primarily surface structure information. Reference is made to gestalt representations of surface acoustic information to derive meanings. Thus, the analyses are quite gross. As the child matures the chunks that can be processed simultaneously at all levels (semantic, syntactic and phonological) increase in duration and, as structural knowledge grows, recursiveness within chunks can be processed and the analysis becomes more detailed or differentiated. The speech signal must be held in mind and represented to allow analysis using whatever structural knowledge is available. It has been suggested that this representation

or categorization of speech is initially acoustic images of morphemes and/or syllables and only later in terms of segments and features of segments. This appears to be the case even though the infant is capable of discriminating between minimally different acoustic features. In summary, the model that appears to be most descriptively adequate is not a "top-down" or "bottom-up" model but, rather an "outside-in" model (Menyuk, 1977).

References

- Cooper, F. (1972): "How language is conveyed by speech", in Language by ear and by eye, J. Kavanagh and I. Mattingly (eds.), 25-46, Cambridge, Massachusetts: MIT Press.
- Eimas, P. (1974): "Linguistic processing of speech by young infants", in Language perspectives: acquisition, retardation and intervention, R. Schiefelbusch and L. Lloyd (eds.), 55-74, Baltimore: University Park Press.
- Fodor, J., T. Bever and M. Garrett (1974): Psychology of Language, New York: McGraw-Hill.
- Garrod, S. and T. Trabasso (1973): "A dual memory information processing interpretation of sentence comprehension", JVLVB 2, 155-167.
- Menyuk, P. (1977): Language and maturation, Cambridge, Massachusetts: MIT Press.
- Menyuk, P. and L. Menn (In press): "Early strategies for the perception and production of words and sounds", in P. Fletcher and M. Garman (eds.), Cambridge, England: Cambridge University Press.
- Newport, E. (1976): "Motherese: the speech of mothers to young children", in Cognitive theory: Vol. II, N. Castellan, D. Pisoni and G. Potts (eds.), Hillside, New Jersey: Lawrence Erlbaum Assoc.
- Savin, H. (1972): "What the child knows about speech when he starts to learn to read", in Language by ear and by eye, J. Kavanagh and I. Mattingly (eds.), 319-326, Cambridge, Massachusetts: MIT Press.
- Staats, A. (1971): "Linguistic-mentalistic theory versus an explanatory S-R learning theory of language development", in Ontogenesis of grammar, D. Slobin (ed.), 103-152, New York: Academic Press.

SOCIAL FACTORS IN SOUND CHANGE: Summary of Moderator's Introduction

Einar Haugen, Department of Linguistics, Harvard University, Cambridge, MA 02138, U.S.A.

The papers offered in this symposium may be divided into "theoretical" and "empirical", even though of course both types of research are represented in all. The papers by Birnbaum, Fónagy, and Malmberg are primarily theoretical, Brink/Lund, Labov, and Peng primarily empirical.

Birnbaum offers for discussion a model of linguistic change originated by Henning Andersen, in which the key word is "abduction", especially applicable to the process of linguistic decoding. Birnbaum is critical of certain aspects of this model, especially its implication that a speech community may be homogeneous or consist of neatly separable generations.

Fónagy is concerned with the idea offered by some that intonation is a non-arbitrary, naturally motivated phenomenon. To disprove this he offers samples from French and Hungarian of how intonations can change their signification over time and become arbitrary expressions associated with particular social groups.

Malmberg takes as his starting point his own earlier studies of the Parisian vowel system, in which he found an "état de langue" which included two systems, a "maximum" and a "minimum" system of vowels between which the speaker could choose. The "minimum" system represented a simplification, which Malmberg attributes primarily to "peripheral" learners of the language, whether they be socially or geographically marginal, i.e. lower class or colonial, the latter exemplified by Spanish in the Americas.

Brink and Lund (here Brink/Lund) have completed a massive study of Copenhagen speech from 1840 to 1955, based on the recorded voices of speakers born between these dates. Their researches have uncovered some 60 phonetic changes (which they call "sound laws") that characterize this period and permit them to classify their speakers into two groups, according to whether they speak "high" or "low" Copenhagen.

Labov's paper sums up some of the conclusions at which he has arrived on the basis of his classic studies of Martha's Vineyard in Massachusetts, the Lower East Side of New York City, and the city of Philadelphia. He has been a pioneer in developing a technique of selecting "social markers" which permit him to place speakers rather accurately on the socioeconomic scale.

Finally, Peng presents a summary of his studies of the linguistic changes in the city of Tsuruoka in Japan, data gathered by his colleague Nomoto in the years 1950 and 1971, in many cases from the same informants. Out of this material he has drawn conclusions that reduce the time span within which one can observe linguistic change even more drastically than Labov: he contends that it is possible to identify linguistic change within a single generation.

Each of these papers brings something to the elucidation of a problem that has baffled linguists ever since the regularity of sound change was firmly established early in the nineteenth century. The causes of sound change were vainly sought in everything from climate to human physiology. Until recently linguists were convinced that change was so slow that it was inaccessible to direct observation. Diachronic linguistics became the study of the past, historical and even paleontological, while a synchronic linguistics sprang up which was based on assumptions of heuristic stability and uniformity, as language might wishfully appear to the prescriptive grammarian.

The Prague School declared that the ideal standard language should possess both stability and elasticity, i.e. it should be flexible enough to change and yet conservative enough to seem unchanging. They did not realize that this paradox could and must apply to every variety of human language; its latest synonym is Labov's expression in describing his concept of language: "orderly heterogeneity". He opposes this to Chomsky's "ideal homogeneity", but in fact his variable rules are a formalization of the concept of "elasticity", while categorial rules reflect "stability". Questions have been raised about the statistical nature of variable rules: how can a speaker know that he is going to use one sound 66% of the time and another the remaining third?

Part of the answer comes from the painstaking analysis by Brink and Lund of the recorded materials from Copenhagen. They have offered no statistics, but in their big book (unfortunately available only in Danish) they have traced from decade to decade how certain changes arose, how speakers vacillated from one to the other form, and how new generations resolved the conflict by choosing one or the other of the alternatives. It is clear that the concept of "choice" with which Malmberg operates has been at work, but it is not clear that it has been a choice between two or more coherent levels of speaking. Even with the masses of data now being accumu-

lated in such studies, including Labov's and Peng's, we are far from knowing why these choices are made, either individually or collectively. Such a study would be an infinite regression going far beyond the realm of linguists' competence, especially if the goal were to construct some kind of predictive model that would tell us what kind of changes the future will bring. Brink/Lund's material shows clearly that at any given point in time there is a great deal of unstructured heterogeneity, vacillation which may either lead to innovation or regression.

Our contributors differ sharply on certain crucial aspects of the problem. Brink/Lund flatly assert that regular "sound changes occur between generations (in our opinion innovations come from children, who - under mutual influence - retain while growing up a few of their originally many deviations from the adult language)". They refer to recordings of the same persons from 30 to 50 years apart in which one could detect virtually no change. Against this Peng claims that the changes passed on are those that young people have developed up to the age of 35, when they communicate them to their children. Against these extreme views we may place Birnbaum's judicious remark that there is a "continuous pattern-setting effect of parents on children, teachers on students, leaders on followers, older on younger playmates and fellow workers, more prestigious on less prestigious...".

There is also some difference of opinion on the role played by social classes and other groups in the activation of change. Labov has found that the upper working or lower middle class leads in changing, while Brink/Lund hold that in general the lower classes of Copenhagen have been in the lead, as being the majority, if not the most prestigious socially. The difference may be more terminological than real, for it is hard to compare the finely graded scale of socioeconomic status developed by Labov with Brink/Lund's linguistic division of the entire population of Copenhagen into two groups, the H-speakers and the L-speakers. On one point all are agreed: that women have more H-features than L-features, though Brink/Lund will not grant that there is a special female sexlect.

Both Peng and Malmberg emphasize that it is not language that changes, but people who change language. This is clear enough when we speak of the adoption of new words or the learning of new dialects and languages, but for phonology the functioning is so automatized and deeply embedded in the subconscious that it has been

difficult to find any clear social causes for specific changes, e.g. Umlaut or the Germanic consonant shift.

I would suggest that we do know a good deal about the causes of sound change, but we have made little progress in predicting its results. But at least we now have techniques and instruments that enable us to catch it on the wing and study it while it is going on. We still have a long way to go before we can learn to control it, if we should ever wish to do so. In this respect we are no worse off than any other social science.

ONGOING SOUND CHANGE AND THE ABDUCTIVE MODEL: SOME SOCIAL
CONSTRAINTS AND IMPLICATIONS

Henrik Birnbaum, University of California, Los Angeles, USA

Underlying the present discussion of some aspects of sound change is the notion that language not only, as energeia, (or, explicitly, as a set of largely automatized processes definable in more or less accurately phrased rules), is susceptible to formal analysis of some degree of descriptive adequacy and explanatory power but that, in addition, it can be conceived of as an inherent and integral part of human thought and imagination. Adopting the latter point of view, language can be said to form a conceptualized (verbalized) mirror image of mental activities (cf. the notion of language as the primary modeling system, elaborated in Soviet semiotics). The former approach, concerned with building models of linguistic structure (or parts thereof), views language as a - particularly sophisticated - semiotic subsystem (operating within the parameters set by its specific neurophysiological premises) and strives to explain its functioning in this capacity. The other kind of inquiry into the nature of verbal communication places the chief emphasis on language as a cultural manifestation of the human mind (in the sense of Geisteswissenschaft) and seeks to understand its performance in society. The former approach may be termed generative (in the broadest meaning), the latter hermeneutic. Both, if applied pragmatically and without any ad hoc constraints, have a sociolinguistic dimension.

It is a fairly common view that sound change takes place gradually in a series of minimal, barely noticeable adjustments and modifications at the phonetic (subphonemic) level and that it is only at the functional or semantically distinctive (phonemic) level of sound production and, in particular, perception that the impression of abrupt sound change obtains.

Some years ago, Andersen (1973), while critical of 'standard' TG phonology but adopting a broadly generative approach to linguistic inquiry in terms of positing specific speaker/hearer 'grammars', i.e., sets of rules generating acceptable sound sequences (utterances), proposed an intriguing model of phonological change. In addition to induction and deduction, he introduced, following Peirce, a third mode of inference termed abduc-

tion. Applying deduction and abduction specifically to sound change, Andersen (1973, 777, fn. 13) points to the "unique role of abduction ... vis-à-vis the other modes of inference, which merely test what has been arrived at by abduction" and suggests that "one can evidently describe the process of encoding as essentially deductive, and that of decoding as abductive". In closing, he submits (1973, 791) that while early structuralism (Jakobson) "could insist only that every phonetic innovation be interpreted in terms of the system that undergoes it ..., it is [now] possible to interpret every phonological innovation - abductive or deductive - in terms of the system that gives rise to it".

In a subsequent paper, Andersen (1974, esp. 25-6, 41), in discussing and summarizing his typologies of innovation in the content and expression systems of language, distinguishes between adaptive and evolutive innovations, with the former subclassified, on the expression plane, into remedial and contact innovations; the evolutive innovations are subdivided into deductive and abductive, with the abductive innovations of the expression plane further specified as pertaining either to the phonemic system (a) feature valuation, b) segmentation, c) ranking), or to pronunciation rules. In a more recent study, with his theoretical reasoning again firmly grounded in Slavic diachronic and dialectal data, Andersen (1978, section 4.2) arrives at the conclusion that we must "acknowledge that conceptual factors take precedence over perceptual or articulatory ones in determining how a phonological system may be changed as it is transmitted from generation to generation ... and recognize that it is the structuring principle of linguistic form - the fact that the speech signal must be segmented, that distinctive features are binary, and that they must be ranked - and not the articulatory or acoustic or perceptual substance that shape its historical development. We are led to conclude that the ultimate source of dialect divergence - and of linguistic change in general - is the process of language acquisition, in which the speakers of a language impose form on the fluctuating and amorphous substance of speech." Novel and incisive though these formulations are, they not only allude to Jakobson's views about DF analysis and language acquisition, but in their reference to form and substance, content and expression also echo some of the basic tenets of glossematic theory. Yet, essential-

ly, the abductive model of sound change, pertinent, above all, to the decoding process, is of course Andersen's, at least as consistently formulated by him and solidly underpinned by theoretical considerations. The model implies that the output of 'grammar 1' serves as the input to 'grammar 2' which in turn yields a reinterpreted 'output 2', slightly, yet significantly different from 'output 1' (1 and 2 here symbolizing successive generations); cf. esp. Andersen (1973), 767 and 778, figs. 1 and 2.

It should be noted, however, that observations and inferences of a similar kind have been made with regard to phonological change also prior to Andersen's sketching of his model of abductive innovation in phonology, as well as after the appearance of his first, seminal paper on the subject. As an example of the latter - arrived at independently, it seems - may be quoted some remarks made by Hetzron in discussing two principles of reconstruction in genetic linguistics. Thus, Hetzron (1976, 96) writes: "In diachrony ... what is transmitted from generation to generation is not the structure, but a set of data which is analyzed by the child acquiring the language so that he could establish a structure for his own use. Language change is precisely justified by the fact that a subsequent generation may analyze the facts perceived by learning the language from the older generation, and this may eventually require some adjustment in the facts, some modification of the perceivable data". To be sure, Hetzron's formulation is less precise than Andersen's in addition to being couched in traditional structuralist ('taxonomic') rather than in broadly generative terms. But in essence, this is in line with Andersen's more elaborate and tightly argued model of phonological innovation.¹

When stating his premises, Andersen (1973, 767) wrote: "What is needed is a model of phonological change which recognizes, on the one hand, that the verbal output of any speaker is determined by the grammar he has internalized, and on the other, that any speaker's internalized grammar is determined by the verbal output from which it has been inferred." And he qualified

(1) For an earlier comment on the similarity of Andersen's and Hetzron's reasoning and a first criticism of a shortcoming they, in my opinion, share, see Birnbaum (1977), 28-30.

his theoretical framework by adding the crucial requirement: "The model that is needed must show how phonological innovations can arise in a homogeneous speech community ..." While the broadly generative (and logic) premise sketched seems most useful indeed, the formulation of the sociolinguistic condition is somewhat questionable (his reference to Labov's definition notwithstanding). What, in fact, is a homogeneous speech community? And what exactly is meant when Andersen (like Hetzron) speaks about the transmitting of a phonological system (or a set of data) from generation to generation? As I had an opportunity to caution (Birnbaum, 1977, 30): "... the transmission of a linguistic system or subsystem (or a grammar or grammatical component generating this system or subsystem) from one generation of speakers to the next must not be conceived of in all too rigid, mechanistic terms since the distinction of successive generations in any real speech community is never very clear-cut and easily ascertainable." Put differently, even though sound change in reality — on the phonetic level, accessible to physical scrutiny and measurement — occurs gradually and it is only on the more abstract phonemic level that one sound, at some point, simply replaces another, it is nonetheless a fact that, given the passage of time, an actual sound shift (e.g., *e* > *o*, *ou* > *u*; *d* > *t*, *k* > *č*) is ascertainable also at the phonetic level. How do such phonological changes come about? Surely not as a result of any simultaneous gradual adaptation by each entire membership of a number of clearly definable consecutive generations. Obviously, a real speech community is never truly homogeneous, nor does it consist of a limited set of neatly separable generations.

Considering the interpenetration of synchrony and diachrony — in phonology, ongoing sound change — it would seem more realistic not to posit a limited set of coexistent generations at any given time (as is implied in Andersen's abductive model as well as in Hetzron's informal reasoning) but rather to assume the continuous pattern-setting effect of parents on children, teachers on students, leaders on followers, older on younger playmates and fellow workers, more prestigious on less prestigious population groups, etc., all interacting at various ages and stages of their development. While such a view of society and language does not vitiate the validity of Andersen's abductive model of sound

change altogether, it certainly makes his scheme more problematic; also, given these complicating factors, his technique for describing, analyzing, and explaining actual phonological innovation is in need of further refinement.

Here one more point should be briefly discussed. It has become customary to attribute great significance to the process of acquiring language, i.e., the mastering of one's native tongue in early childhood, also when it comes to explaining certain basic facets of sound change. (The partial or complete acquisition of a foreign language presents analogous but also additional problems.) Andersen's abductive model, in this respect influenced by Jakobson's work on child language, is but one example of this conception. However, it seems worth considering whether, precisely as regards modifying one's pronunciation habits, i.e., introducing incipient or, occasionally, even full-fledged phonological innovations, it is actually in early childhood (say, before the completion of the fifth year) that the definitive articulatory profile of a person is usually formed and stabilized. Rather, I would submit, that is the age when growing-up speakers, by imitating their elders, attain the same or nearly same pronunciation as their models. True, in the process they may very well, by 'misreading' (i.e., slightly incorrectly perceiving) the phonetic output of 'grammar 1', internalize, initially, at least, a somewhat deviant 'grammar 2' (or, rather, its phonological component) producing — following Andersen's reasoning — a phonetic 'output 2' not fully identical with 'output 1' of their model. Yet, very often (if not as a rule) most of the misperceived pronunciation is subsequently noticed and rectified except, perhaps, where the resulting differences in pronunciation are so minimal as to be considered insignificant even by the maturing child; it is only their cumulative effect over a longer period of time that ultimately may give rise to a genuine sound change. However, it appears that attitudes at a somewhat older age, especially in the teens, may more directly, noticeably, and lastingly affect pronunciation habits and cause partial or even full sound shifts (or, rather, sound substitutions) to occur within one generation. I am referring here to the fashionable pronunciation or talking fads which, particularly in our day and age, so markedly leave their imprint on the speech habits of the teenage generation. It

is my impression, based on observations from several languages, that the modification of the articulatory manners and preferences affecting these young people are more radical, since they are deliberate, than are the difficulties in imitation and pronunciation adjustment encountered in early childhood. If Andersen's abductive model of phonological innovation is to be applicable also to currently observable sound change – and not only to interpreting and elucidating instances of historically attested or reconstructed phonological shifts – these sociolinguistic and psycholinguistic considerations will somehow have to be accounted for in his model.

Viewing sound change primarily as a sociolinguistic phenomenon, best studied while in progress, it must be said – with all due respect to Labov's 'integrated' explanation² – that we are still far from genuinely and fully grasping its causes. So far, there has not been much more than a general realization of the permanent and highly creative interplay between, on the one hand, language's striving for economizing (ultimately tending toward ellipsis while preserving a measure of redundancy as a safety valve to ensure comprehension and information transfer; cf. Martinet 1955) and, on the other, its making for diversity of expression to distinguish among even the finest shades of meaning. Though sound, at the phonemic level, does not by itself carry, but merely distinguishes meaning, it and its modification are crucially affected by this dialectic tension characteristic of language as a semiotic system.

(2) The study of ongoing sound change viewed in its social setting has in America been pursued, in particular, by Labov; cf. esp. Labov (1963), (1966), (1970), (1972), (1973); and Labov et al. (1968), (1972); for a brief assessment of Labov (1973), see, e.g., Birnbaum (1975), 284-6. Of more recent work by scholars with other ideas, see, e.g., Bailey (1973), Peng (1976), and Itkonen (1977).

References

- Andersen, H. (1973): "Abductive and Deductive Change", *Lg.* 49, 765-793.
- Andersen, H. (1974): "Towards a Typology of Change: Bifurcating Changes and Binary Relations", in: *Proceedings of the First International Conference on Historical Linguistics* (J. M. Anderson & C. Jones, eds.), II, Amsterdam & Oxford: North-Holland, 17-60.
- Andersen, H. (1978): "Perceptual and Conceptual Factors in Abductive Innovations", in: *Recent Developments in Historical Phonology* (J. Fisiak, ed.), The Hague: Mouton [in press].
- Bailey, C.-J. N. (1973): *Variation and Linguistic Theory*, Arlington, Va.: Center for Applied Linguistics.
- Birnbaum, H. (1975): "Typological, Genetic, and Areal Linguistics: An Assessment of the State of the Art in the 1970s", *FoundLg* 13, 267-291.
- Birnbaum, H. (1977): *Linguistic Reconstruction: Its Potentials and Limitations in New Perspective*, Washington, D. C.: Institute for the Study of Man (*The Journal of Indo-European Studies*, Monograph No. 2).
- Hetzron, R. (1976): "Two Principles of Genetic Reconstruction", *Lingua* 38, 89-108.
- Itkonen, E. (1977): "The Relation Between Grammar and Sociolinguistics", *Forum Linguisticum* 1:3, 238-254.
- Labov, W. (1963): "The Social Motivation of a Sound Change", *Word* 19, 273-309.
- Labov, W. (1966): *The Social Stratification of English in New York City*, Washington, D. C.: Center for Applied Linguistics.
- Labov, W. (1970): "The Study of Language in Its Social Context", *Studium Generale* 23, 30-87.
- Labov, W. (1972): *Sociolinguistic Patterns*, Philadelphia: University of Pennsylvania Press.
- Labov, W. (1973): "The Social Setting of Linguistic Change", in: *Current Trends in Linguistics* (T. A. Sebeok, ed.), 11, The Hague & Paris: Mouton, 195-251.
- Labov, W. et al. (1968): "Empirical Foundations for a Theory of Language Change" (with U. Weinreich & M. I. Herzog), in: *Directions for Historical Linguistics: A Symposium* (W. P. Lehmann & Y. Malkiel, eds.), Austin & London: University of Texas Press, 95-188.
- Labov W. et al. (1972): *A Quantitative Study of Sound Change in Progress*, 2 vols. (with M. Yaeger & R. Steiner), Philadelphia, Pa.: U. S. Regional Survey (NSF GS-3287).
- Martinet, A. (1955): *Économie des changements phonétiques*, Berne: Francke.
- Peng, F. C. C. (1976): "A New Explanation of Language Change: The Sociolinguistic Approach", *Forum Linguisticum* 1:1, 67-94.

SOCIAL FACTORS IN THE SOUND CHANGES OF MODERN DANISH

Lars Brink and Jørn Lund, University of Copenhagen, Denmark

In the following we will present some of the major results of our research¹ on the role of social factors in the sound changes of Modern Danish. Our investigations deal with phonetic history based on phonetic sources. The oldest living informants we tested were born in 1875. Edison and others made it possible to investigate an added generation, namely a solid group of informants born from 1840 on. The oldest Danish voice preserved, to our knowledge, must have resounded for the first time in 1813, two years before Waterloo! The recordings comprise at least 10 informants per 5-year period, except for the very first years. Our goal was to register and describe all ascertainable pronunciation changes for those informants raised in Copenhagen and to survey pronunciation and its development outside Copenhagen.

The phonetic history of Copenhagen speech in the period treated reveals an amazing wealth of sound changes. This is largely the result of ca. 60 regular sound laws. We have described the sound changes by consistently referring to the birth date of the informants - since this was our first significant result: When we arranged the material according to the informants' birth dates, numerous sharp boundaries appeared. In fact, often a clear shift between informants born in two subsequent 5-year periods became apparent. If we arranged the material according to recording dates rather than birth dates, no changes could be demonstrated at all, unless the age distribution in the groups was kept painstakingly constant. And even then, the sound changes would only appear as weaker shifts in a sea of variation. - The reason, of course, is that regular sound changes occur between generations (in our opinion innovations

(1) "Dansk Rigsmål I-II. Lydudviklingen siden 1840 med særligt henblik på sociolekterne i København" (Standard Danish I-II. The phonetic development since 1840 with special regard to the sociolects in Copenhagen). 823 pp., Gyldendal 1975; "Udtaleforskelle i Danmark" (Differences in pronunciation in Denmark). 113 pp., Gjel-lerup 1974; "Regionalsprogsstudier" (Studies in regional language), in Danske Studier 1977 (by Jørn Lund); "Om lydlove" (On sound laws), paper presented to the Soc. of Nordic Phil. in Denmark 1977 (by Lars Brink); and an investigation of the linguistic situation in Copenh., 219 pp., 1978 (primarily by Jens Normann Jørgensen). - Here we economize with respect to examples and documentation, and for definitions and methodology we refer to the above works.

come from children, who - under mutual influence - retain while growing up a few of their originally many deviations from the adults' language), and that influence from younger speakers on older ones is modest and never strong enough to allow the older to "catch up" with the younger. We have several recordings of the same speakers made as far as 50 years apart, and they reveal only slight differences. On the whole, it is our experience that most adults who do not change their milieu are only weakly influenced phonetically by the next generation.

It was clear from the start that the material had to be arranged according to sociolects. Certain linguistic features are correlated with high social status (in a certain generation, in a certain area), i.e. the feature becomes more and more common as we progress up the social ladder, e.g. the pronunciation [ʃo⁺'fø+g'] (chauffeur). Others are correlated with low social status, e.g. [ʒæ⁺'fø+g'] (here we employ IPA with unmodified values), and still others are socially neutral. Thus, we could arrive at two, and only two, sociolects in Copenhagen: High and Low Copenhagen, i.e. the languages with (almost) totally high - resp. low - or neutral linguistic features, and, of course, intermediary forms. In the following we will trace the major trends in the development of the two Copenh. sociolects.

Prior to ca. 1750 (birth year), there were practically no social linguistic variations. There were a few differences in the pronunciation of foreign words, and certain folk etymologies must have belonged solely to the lower level. The linguists of the day operate with many divisions: Copenh./provincial, the various dialects, free speech/reading pronunciation, but no one ridicules the common man's deviations from the learned except for certain "distortions" of foreign words. A number of somewhat later authors write, on the contrary, that pronunciations, inflections, etc. which clearly belong to craftsmen and servants in their day were quite common among the learned in previous generations. This agrees with our findings that the numerous certain L-features (L- = low-) in the 1800's can nearly always be traced back to a time when they were common in higher circles. The situation was the same in the country districts. The linguistic interaction in these less specialized societies was probably simply too great to allow social characteristics to arise. A town like Copenhagen

was small (in the 1700's not quite 3 km²) and densely populated. The division into better and poorer neighbourhoods belongs to a much later age.

The social uniformity refers to language, not to all aspects of speech,² since speech involves many language-independent aspects. Of course, there will always be statistical differences between high and low speech. Due to statistical differences in interests, experiences, intellectual equipment, etc., there will be statistical differences e.g. in conversation topics, sentence length, metaphors, irony, slips of the tongue, syntactic anacolutha, etc. But the fact that L-speakers might discuss boxing more often, or perhaps employ more anacolutha than H-speakers has nothing to do with their language, i.e., the traditionally transmitted set of rules which govern their speech. No linguistic rule recommends or forbids talk of boxing, and it is self-contradictory to maintain that L-language could be viewed as requiring the use of anacolutha, since by this we naturally do not mean deviations in regular speech from regular writing but syntactic deviations in actual speech from regular speech (speech which does not in the least violate the linguistic rules of the speaker). For everyone, their number is great.

The recognition of a social uniformity prior to 1750 must be modified on a few points. The various work spheres have always had a set of terms generally unknown outside the field. This is so obvious that we do not consider it in the notion of a sociolect. If we did there would in every society with a division of labour be just as many sociolects as fields. Other differences in vocabulary involve literary words and learned, foreign words. Finally, in the lower circles there must have been somewhat weaker taboos on swear words and obscene words. None of the two last mentioned situations involve actual sociolect differences: They are even on the highest or lowest social level extremely individually determined. True sociolect features do not reflect individual personality features of the speaker but merely the habits of his surroundings. The decisive argument is that the above mentioned

linguistic features correlate to a higher degree with characteristics such as certain types of knowledge or attitudes toward taboos and only indirectly, and more weakly, with social class. In many cases these non-linguistic or non-sociolectal situations are mixed together with the true sociolect features such that one can obtain the quite distorted impression that high and low language are arranged on a scale, the extremities of which are written academic language and children's speech: Written academic language contains a maximum of regularity, literary terms and urbane expressions - children's speech a minimum.

In the time following, a long series of true sociolect features emerges in pronunciation, inflection, syntax and the core vocabulary. In the beginning, ca. 1750-1800 (birth year), it is a matter of a few features, always such that the lower social levels retain an older form while the higher levels take on a new form or limit themselves to one of two older double-forms. No written source indicates the existence of sound changes of the type: the uneducated have zero where the educated have [h], etc.

Not until ca. 1800 (birth year) do we find this type of differences. The oldest socially staggered sound changes we can ascertain from the phonetic material from 1840 on must have been initiated at the earliest in the generation born in 1800. The first indication in written linguistic sources of such differences we find as late as in the 1880's among authors born after 1850. Apparently no one was aware of these differences for a long time.

From ca. 1900 (birth year) the new changes do not appear nearly as staggered socially, and a series of new sound changes originating in the previous period, both in the L- and in particular in the H-languages, now become reversed. Among the youngest living adults there are still differences left, but on the one hand, these show a tendency toward leveling, and on the other hand, the strength of the social correlations themselves is decreasing. This is all probably due to greater social mobility and integration, and of such individual factors we feel, without being able to prove it, that the most important cause is the fact that from ca. 1900 it became common for the educated to send their children to free, state-supported schools. Radio and television have had no influence, since leveling has most often been in the direction of the L-language, and since in radio and television up to ca. 1970 the H-language was almost always used.

 (2) in the basic sense "the actual output of speaking" - including all its phonetic and non-phonetic sides.

Of the many regular sound changes appearing after 1800, the majority originate in L-Copenh., e.g. [aɨ > ε] before alveolars and zero, [a-: > ε:], the numerous vowel openings before and after r, [ɣh > ɣ^{sh} > ɣs], and the openings of the medial and final spirants [-v > -v̥] [-y > -wɾ/-jɾ] ([wɾ] after back, [jɾ] after front vowels), [-ðɾ > -jɾ] and [-v > -wɾ]. It may seem surprising that the H-sociolect does not lead the way, but this is only at first glance. First of all, it is only natural for new changes to arise in the largest linguistic community. Secondly: for a long time the change is not obvious to speakers, characterized merely by the fact that a growing number of speakers begin to show uncertainty with respect to the new and the old form. The change usually spreads to the H-sociolect one or two generations later where the same situation of uncertainty then appears, and first then do attentive H-speakers become aware of it and, normally, offended, but then it is too late, since the remaining younger H-speakers are pushed from three corners: from the majority of L-speakers, from the minority of H-speakers, and, most importantly, from the "inherent plus-value" of the new forms. It is namely no accident that the new forms could expand from non-existence and thus defy the general imitation tendency (!). - A number of changes, however, originate in the H-language, e.g. the shift of back vowels: [ɔ+ > ɔɨ+], [ɔ+ : > ɔɨ+ :], [ɔɾ > vɨ], [a+ > aɨ], and vowel shortening before vocoids, especially before [ɣ], and some changes show no clear social staggering.

Sociolect and sex. The two sociolects show uneven distribution with respect to sex. Women generally have more H-features than men. Indeed, the group of pure L-Copenh. speakers is made up of more than twice as many men as women; with respect to the pure H-speakers, the difference is not nearly as great. However, we are still dealing with sociolects and not sexolects: the difference between high and low is in every instance investigated greater than the difference between women and men. This can hardly be due to anything but the fact that women for some reason have greater social aspirations than men and are perhaps also more attentive, but these factors need not be the main reason in every concrete instance in which, e.g., the girls in a group of siblings have more H-features than their brothers, since once the sexual difference has come to exist, it can then largely be maintained simply by the fact that girls are more apt to imitate other girls and women.

Orthographic influence. Our investigation shows that regular sound changes, not orthography, are the most important sources of phonetic change. The ca. 60 sound laws which we have ascertained are either neutral towards the orthography or, the majority, directly contrary to it. They have been accompanied by numerous new mergers in vocabulary, e.g. 'ret' = 'rat' (right, steering wheel): [ʰaɨɣ], 'sagn' = 'savn' (legend, lack): [ʰsaɨwɾ'n], 'æder' = 'edder' (eats, venom): [ʰeɾðɾ'vɨ], 'lure' = 'luer' (eavesdrop, flames): [ʰlu+ɾ], and 'løjet' = 'lodde' (lied, solder): [ʰvɨðɾ'ðɾ], and in general this has made spelling more difficult. In addition, many changes in isolated words are contrary to the orthography. We have found many instances of orthographic influence, in proper names, in less common words, and in foreign words in the L-language and also some cases in the core vocabulary; but in ordinary spoken language, in a running text, orthographic influence accounts for a very small portion of all the changes in pronunciation occurred between the 1840 generation and today's youth.

The range of the standard language. In Denmark there exists a non-localizable standard language or rather a complete set of un-localizable linguistic forms, i.e. forms which are not tied to speakers raised in particular areas. By comparison with dialects we have attempted to show that these non-localizable forms, wherever they contrast with the respective dialects, have their historical origin in Copenh. speech. A non-localizable standard language is first completely established when non-localizable forms exist without exception. First with the generation born around 1825 was the last "resistance" to Copenh. forms abandoned, namely when the Copenh. [a+jɾ] and [ɔɾjɾ] diphthongs made their way into the provinces. But just because a complete non-localizable language is available it is not certain that anyone speaks it in its purest form. Actually, and quite naturally, only inhabitants of Copenhagen did so. Not until the generation born in 1880 do we find such informants raised in the provinces, namely H-speakers from the Sjælland market towns. Outside of Sjælland all thoroughly investigated informants possessed certain (but not the same) local features. The local characteristic increases everywhere in the market towns as the social status falls, but even on the lowest social levels the language, today, is much closer to Copenh. speech than to the locality's original dialect, now relegated to the rural

areas and dying out even there. On Sjælland we know of no true dialect speaker born after 1920. - Even the newer developments in Copenh. speech, including the many originating in L-Copenh., spread to the entire country, always somewhat "delayed". Thus, the provinces are characterized both by local forms and by a number of standard-archaisms, archaisms which in the end can become locally bound, namely once the new form has been adopted in (a portion of) the rest of the country. - There exist other centers of development than Copenhagen. In fact, all larger province cities exert an influence on their surroundings. For this reason certain vigorous local forms can now be found in larger provincial areas outside their original dialect area, but primarily these centers function as mediators for the Copenh. features which always influence the larger cities earlier and more forcefully than the smaller.

The Copenh. forms in general, as widespread as most of them are now, are no longer felt by speakers to be Copenh. forms, which they actually are only historically. Their success is thus not due to any capital city prestige. This factor was no doubt important in previous centuries when the non-localizable language was in the process of being established. In the provinces there must exist a certain general high-social atmosphere around them, but not even this factor can be the decisive one, since L-Copenh. forms, as mentioned, also spread energetically. We have attempted to show that logically the language of a capital city can succeed in spreading its forms to the rest of the country purely by contagion, namely according to what we term the Napoleon-principle: the enemy is slain where he is weakest and immediately enrolled in the victor's troops. But, of course, prestige plays a significant role. We feel, however, that the most important emotional attitude toward linguistic forms is one of egocentricity, i.e. preference for habitual forms; spot-checks have shown that most L-speakers actually prefer their own forms to H-forms and provincial-speakers their own provincialisms to standard forms (surely, H-speakers' preference for H-forms is greater, since here habit and prestige-value pull together). But the provincial speakers can naturally get their egocentric resistance shattered under sufficiently massive bombardment with standard forms. - Above all, it must be rejected that the orthography or radio and television have created the standard language. The orthography does not indicate the basic quality of the

sounds; the fact that the standard language possesses the original Copenh. qualities [b, d, g, v, w, t, e, r, o, i:], etc., corresponding to written b-, d-, g-, r-, -d, v-, -v, æ, o, no soul can read from the written picture. Nor does the orthography indicate stød, stress or pitch, and regarding the distribution of sounds the Danish orthography is so archaic that it agrees frequently and significantly with other dialects better than with Copenh. speech (e.g. in the case of -p-, -t-, -k-, hv/v, hj/j, nd/nn, ld/ll, weakly stressed -et; -eg-, -øg-, i, y, u + nasal, geminates). To be sure, the orthography can delay the spread of contradictory Copenh. forms, but it has not been able to stop them. The influence of radio and television has to be modest. The developments mentioned above were well under way (the dialects being long since extinct in all larger market towns, in Aalborg, for instance, already with the generation born around 1800) when radio broadcasting began in 1926, and if this factor was significant today, the degree of local characteristics would not reflect so strongly in part the distance from Copenhagen and in part the degree of urbanization (radio and television consumption does not run parallel to these factors).

Language is a social phenomenon. But language is such a varied activity that it also possesses other aspects. The numerous Copenh. sound changes we have encountered are naturally not the result of normal interaction and imitation, nor of imitation combined with prestige, since the new forms initially have no prestige, they expand from zero-level, and no one realizes initially that he has a new pronunciation, and even less where it comes from. New pronunciations must of necessity possess an inherent plus value in order to progress victoriously. Of course, imitation is one element, but the very dynamics of the development involve the fact that the new pronunciation is initially adopted and adhered to by speakers who far more often hear the old one. The plus value may be connected with errors in perception or production, originating independently with many children (or, theoretically, with one) or with an easier articulation. In other words, when we speak of true sound changes, not just old (standard) forms replacing other old forms in other areas through borrowing, we must not forget that sound change is essentially a non-social phenomenon.

STRUCTURE ET ASPECTS SOCIAUX DES CHANGEMENTS PROSODIQUES

Ivan Fónagy, C.N.R.S. Paris

1. Nous ne disposons à l'heure actuelle d'aucune description d'un changement prosodique ayant eu lieu dans le passé. Cette absence totale de témoignage a pu être interprétée comme indication, sinon comme une preuve ex silentio de l'immuabilité des formes d'intonation. Elise Richter (1933) attribue la stabilité des formes mélodiques à leur caractère naturel, motivé. Or, rien que la diversité de l'intonation des langues romanes, ou d'autres langues appartenant à la même famille, aurait pu éveiller des doutes au sujet d'un principe de l'immuabilité intonative. Des observations récentes semblent indiquer que les structures mélodiques, vues sous l'angle de la diachronie, sont tout aussi mobiles qu'à l'intérieur d'un système linguistique à un moment donné. Les contradictions qui reflètent, dans les cadres de la synchronie, les changements de phonétique segmentale (Doroszewsky 1935, Fónagy 1956, Labov 1972) caractérisent également la prosodie dans différentes langues.

2. Le long débat, parfois très animé (Gill 1936, Fouché [1936] 1952, 49), qu'a provoqué l'ambiguïté de l'accent en français moderne à partir de la deuxième moitié du siècle dernier (Paris 1862, Passy 1891, Meyer-Lübke 1890) peut être interprété comme une "dramatisation" des contradictions inhérentes au système accentuel. L'accent frappe la dernière et/ou la première syllabe des groupes accentuels en fonction des contraintes segmentales rythmiques, syntaxiques, sémantiques, et surtout, selon le genre du discours, les habitudes professionnelles ou individuelles du locuteur. L'accentuation se présente comme une fonction à variables multiples. Le nombre, l'importance et la valeur de ces variables changent continuellement, et ceci depuis le début du dix-neuvième siècle selon le témoignage des grammairiens (Scoppa 1816, 220).

Toutes conditions égales par ailleurs, il est plus probable que l'accent frappe la première syllabe de l'unité accentuelle si

- a) cette syllabe est fermée;
- b) sa voyelle est susceptible d'être allongée;
- c) le premier mot est un déterminant suivi d'un déterminé - surtout s'il s'agit d'adjectifs numériques;
- d) le premier mot (pronom, adjectif, adverbe interrogatif) figure en tête d'une question partielle;
- e) le mot figure en tête d'une phrase impérative;
- f) ou en tête d'une réponse composée d'un seul mot, etc.

La probabilité de l'accent change, toutefois, radicalement d'un genre du discours à l'autre. L'accent barytonique devient sensiblement plus fréquent dans le discours politique (Duez 1978); la distance moyenne entre syllabes accentuées diminue d'une façon drastique dans la présentation des informations où l'accentuation des mots "enclitiques" est presque érigée en règle. Si la probabilité d'un accent frappant un mot enclitique (préposition, article, conjonction, pronom "atone", verbe auxiliaire) est de 0.03 dans le récit (conte de fées), elle varie entre 0.03 et 0.08 dans la conversation et monte à 0.42 dans les informations télévisées (I. et J. Fónagy 1976).

L'extrême diversité de la distribution de l'accent dans l'énoncé reflète un changement en cours et le masque en même temps. L'accent barytonique apparaît comme l'expression d'un contenu mental (émotion, emphase), comme trait caractéristique d'un genre du discours ou d'un style professionnel, et non pas comme une manifestation d'un changement prosodique. Le rapport de cause à effet est interprété en termes de fin et de moyens. Il est intéressant de voir que même des linguistes distingués n'échappent pas à cette vision finaliste propre aux membres de la communauté linguistique (cf. Fouché [1936] 1952, 51 et s., Fónagy 1979, 180).

Les contradictions socio-phonétiques qui semblent caractériser les changements phonétiques en cours sont l'indice du changement d'accent, mais ne fournissent pas une preuve suffisante du changement. Les premiers témoignages sûrs d'une accentuation barytonique datent du début du dix-neuvième siècle (Fónagy 1979, 168 et ss.). Il faut noter également que les "irrégularités" prosodiques relevées par Richard Strauss chez Debussy (v. ses échanges de lettres avec Romain Rolland), ne se rencontrent jamais chez Lully ou Rameau. Nous avons pu comparer les résultats de tests de perception faits à partir d'enregistrements de 1914-1915 (discours de Poincaré, de Deschanel et de Viviani), d'une part, et d'enregistrements de discours politiques prononcés en 1974, d'autre part. Les mots perçus avec un accent principal ou secondaire sur la première syllabe (sans autres accents), ou ceux ayant un accent principal sur la première et un accent secondaire sur la dernière syllabe, sont nettement plus fréquents dans les discours de 1974, malgré les divergences individuelles à l'intérieur des deux groupes. Cet écart est statistiquement très significatif ($\chi^2 = 2357.91$, $p < 0.001$). On ob-

tient un écart plus faible, mais toujours très significatif ($\chi^2 = 19.72$, $p < 0.001$) en comparant au discours de Poincaré une lecture contemporaine du texte, lecture voulue neutre et qualifiée comme telle par cinq juges. (La valeur moyenne de l'emphase attribuée à la lecture était de 0.85 à partir d'une échelle sémantique de sept degrés, 0 - 7.)

Il n'y a pas de changement évident, par contre, dans la densité, la distribution de l'accent au long de l'axe du temps.

Peu d'indication d'un changement de la fréquence de l'accent en fonction des catégories de mots, sauf: le nombre sensiblement plus élevé des enclitiques accentués dans le corpus de 1974 (cet écart est statistiquement significatif, $\chi^2 = 330.07$, $p < 0.001$).

3. Quant aux changements d'intonation en cours, j'ai pu relever au cours des années cinquante l'apparition récurrente de l'intonation interrogative dans les phrases impératives de sujets hongrois. Cette substitution, cette métaphore mélodique, était particulièrement fréquente dans certains groupes professionnels (contrôleurs de tramway, employés de magasin) d'une part, et chez les jeunes d'autre part (Fónagy 1969). Le transfert était à l'époque motivé, c'est-à-dire limité à des cas où l'intonation montante-descendante des phrases impératives impliquait la présence de certains éléments sémantiques de la modalité interrogative, du moins dans la parole des adultes. Les tests de perception qui ont permis de corroborer ce jugement intuitif montraient en même temps que, pour les plus jeunes, l'intonation interrogative s'opposait comme invitation polie à l'impératif proprement dit, sans avoir nécessairement une implication interrogative. Le transfert intonatif était donc plus fréquent chez les jeunes et en même temps moins précis, moins marqué du point de vue sémantique. Ces divergences dans la fréquence et dans l'interprétation du transfert mélodique selon les générations semblaient indiquer qu'il s'agissait d'une métaphore mélodique figée, donc d'un changement d'intonation en cours.

Les enquêtes récentes faites à partir du même corpus montrent que le changement est en nette progression. Chez les jeunes, l'intonation montante-descendante apparaît comme la forme impérative non marquée qui s'oppose à l'ordre énergique, agressif.

Ces conclusions provisoires sont basées sur des tests faits à partir de variantes synthétisées de la phrase Figyelj ide "Ecoute-moi" (litt.: "Ecoute ici"). Ces enquêtes ne sont pas ter-

minées à l'heure actuelle. J'espère pouvoir en présenter les résultats numériques définitifs au cours du Congrès.

L'intonation terminale montante des phrases assertives dans les dialectes de l'est de la Norvège, signalée par Bertil Malmberg (1966), peut être considérée également comme une métaphore figée. Elle est en effet interprétée comme telle par Kloster Jensen. Selon lui, cette forme d'intonation n'est pas primitive dans ces dialectes: "Elle représente la généralisation d'un type à l'origine stylistique, utilisé pour engager l'interlocuteur, tout comme le nicht wahr? allemand ou le n'est-ce-pas? français." (Je le cite d'après Malmberg 1966, 106).

L'analogie entre métaphore lexicale et métaphore intonative vaut donc également pour la diachronie: dans les deux cas, le transfert aboutit à un changement. Le fait du changement reste inaperçu dans un premier temps à cause de la motivation sémantique du transfert, dans un deuxième temps à la suite de la démotivation qui détache la nouvelle expression de sa base sémantique originale et efface, par là, les traces du changement.

4. Il y a une autre forme de changement mélodique qui ne présente aucune analogie avec d'autres changements linguistiques: c'est l'interférence mélodique, l'union de deux formes d'intonation. Ainsi, une configuration mélodique réunit en français moderne l'intonation interrogative et assertive. Nous avons enregistré, au cours d'un "jeu des portraits" et dans des films policiers, une intonation interrogative-assertive qui peut être considérée comme une forme remaniée de l'intonation déclarative. Elle se distingue de l'intonation interrogative par une chute mélodique finale brusque et de l'intonation assertive par une montée rapide dans la syllabe accentuée, qui est cette fois l'avant-dernière (Fónagy et Bérard 1973). Un phénomène analogue se produit en hongrois où une forme mixte réunit l'intonation interrogative et celle de la protestation indignée (Fónagy 1965). La question incrédule épouse une forme triangulaire en français moderne:

Il e^es^t là^à ?!

Elle présente, comme la question incrédule américaine (Hadding-Koch et Studdert-Kennedy 1965), allemande, tchèque (Romportl 1973,

153) ou hongroise (Fónagy 1965), une reprise caricaturale de l'affirmation catégorique et interfère en même temps avec l'intonation interrogative (montée dans la syllabe accentuée). Toutes ces formes complexes sont également distribuées dans l'espace socio-culturel; il n'y a donc aucun indice d'un changement en cours. Le changement a dû s'accomplir à une date antérieure. On comprend mieux, dans cette perspective, la coexistence de deux intonations interrogatives en russe (cf. Romportl 1973, 159 et ss)

a) montée et descente dans la syllabe accentuée, vs.

b) montée finale.

La première (a), qui correspond à la courbe mélodique de la question incrédule de l'anglais, du français, etc., figure aujourd'hui comme l'intonation interrogative non marquée (Bryzgunova 1963), et c'est la deuxième (b) qui doit être considérée comme marquée (exprimant l'étonnement ou ayant une valeur d'évocation). Boyanus (1936), qui selon Romportl (1973, 158) était le premier à signaler l'intonation (a), présente la configuration (b) comme modèle de l'intonation interrogative. Il est probable que la configuration (a) était au début, en russe comme dans d'autres langues, une métaphore mélodique remaniée, mais qu'elle s'est généralisée par la suite en se substituant à l'ancienne forme non marquée.

5. Reste à signaler un type de changement mélodique qui suppose une mutation fonctionnelle des formes d'intonation. Une intonation expressive, suggérant une attitude déterminée, devient particulièrement fréquente dans la parole d'un groupe social et finit par se détacher de l'attitude qu'elle est censée exprimer. L'intonation correspondant à une attitude désabusée (une moue, un haussement d'épaules) frappe par sa récurrence tenace dès qu'on écoute les présentatrices de la Radio-Télévision française dans des contextes qui excluent une telle attitude. Nous avons présenté à deux groupes d'étudiants en linguistique (groupe a, groupe b) quatre échantillons de cette forme mélodique, d'abord isolés du contexte (groupe a) puis en contexte (groupe b). Présentés isolément, ces échantillons ont suggéré un air désabusé, voire ironique. Présentées dans le contexte, les mêmes intonations paraissaient à la plupart des sujets comme neutres et ils les attribuaient à une présentatrice de la télévision (I. et J. Fónagy 1976). Ceci revient à dire qu'une forme mélodique expressive est devenue neutre à l'intérieur d'un

groupe professionnel et a acquis par là une valeur évocatrice (Bally 1921, I 203-349). Pierre Léon (1971, 54 et s.), après avoir mis en évidence le caractère irréversible des changements de fonction (expressive → évocatrice, jamais l'inverse), cite comme exemple le style publicitaire composé des traits distinctifs de l'insistance et de la joie. Les formes d'intonation caractéristiques de différents groupes professionnels hongrois (Fónagy et Magdics 1963) se révèlent, dans la plupart des cas, comme des formes d'intonation émotives généralisées, neutralisées.

La généralisation d'une forme peut dépasser tel ou tel groupe professionnel. La montée terminale exprimait en hongrois, et exprime toujours dans certains contextes, certaines attitudes déterminées (surtout l'attitude justificative ou l'expression de l'évidence). Elle est toutefois moins expressive dans la parole féminine où elle prédomine. Elle est aussi en corrélation avec l'âge des sujets et plus fréquente chez les jeunes (figure 1). Au lieu d'exprimer telle ou telle attitude, ou d'évoquer tel ou tel groupe professionnel, elle acquiert un caractère féminin selon des tests de perception (Fónagy et Magdics 1963, 8), moins dans l'opinion des jeunes que dans celle de plus âgés.

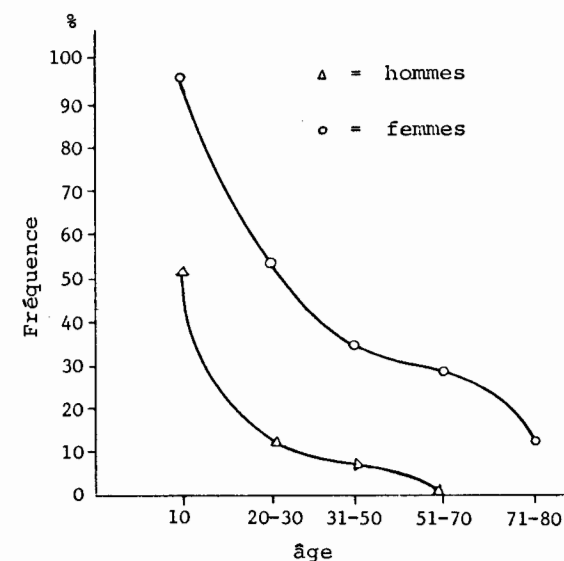


Figure 1

Fréquence de la montée finale dans les questions partielles selon le sexe et l'âge des locuteurs.

6. C'est à force de tels transferts et de telles mutations fonctionnelles que les formes d'intonation de langues apparentées deviennent de plus en plus divergentes d'une langue à l'autre, et que le rapport entre signifiant et signifié devient de plus en plus arbitraire, malgré le lien naturel qui lie l'intonation aux attitudes exprimées.

Résumé

Les structures prosodiques, accentuelles et intonatives, sont loin d'être immuables. Leur changement se reflète sur le plan synchronique par la présence de règles contradictoires. L'intonation change par le transfert, l'interférence et la mutation fonctionnelle des formes mélodiques.

Références

- Bally, Ch. (1921): Traité de stylistique française I-II, Paris: Klincksieck.
- Boyanus, S.C. (1936): "The main types of Russian intonation", Proc. Phon. 2, 110-113.
- Bryzgunova, E.A. (1963): Praktičeskaja fonetika i intonacija ruskogo jazyka, Moskva: Izd. Mosk. Univ.
- Doroszewsky, W. (1935): "Pour une représentation statistique des isoglosses", Bull. Soc. Ling. Paris 36.
- Duez, D. (1978): Essai sur la prosodie du discours politique, Thèse, Université de Paris.
- Fónagy, I. (1956): "Über den Verlauf des Lautwandels", Acta Ling. Hung. 6, 173-278.
- Fónagy, I. (1965): "Zur Gliederung der Satzmelodie", Proc. Phon. 5, 281-286.
- Fónagy, I. (1969): "Métaphores d'intonation et changements d'intonation", Bull. Soc. Ling. Paris 64, 22-42.
- Fónagy, I. et Bérard, E. (1973): "Questions totales simples et implicatives", Studia Phon. 8, 53-98.
- Fónagy, I. et J. (1976): "Prosodie professionnelle et changements prosodiques", Français Mod. 44, 193-228.
- Fónagy, I. et K. Magdics (1963): "Das Paradoxon der Sprechmelodie", Ural-Altäische Jb. 35, 1-55.
- Fónagy, I. (1979): "L'accent français: accent probabilitaire", Studia Phon. 14, sous presse.
- Fouché, P. (1952): Etat actuel du phonétisme français [1936]; introduction à la phonétique historique du français, Paris: Klincksieck.
- Gill, A. (1936): "Remarques sur l'accent tonique en français contemporain", Français Mod. 4, 311-318.
- Hadding-Koch, K. et M. Studdert-Kennedy (1965): "Intonation contours evaluated by American and Swedish listeners", Proc. Phon. 5, 326-331.
- Labov, W. (1972): Sociolinguistic patterns, Philadelphia: Univ. Press.
- Léon, P. (1971): "Essais de phonostylistique", Studia Phon. 4, Montréal: Didier.
- Malmberg, B. (1966): "Analyse des faits prosodiques - problèmes et méthodes", Cahiers de ling. théor. appl. 3, 99-108.
- Meyer-Lübke, W. (1890): Grammatik der romanischen Sprachen I, Lautlehre, Leipzig: Reisland.
- Paris, G. (1862): Etude sur le rôle de l'accent latin dans la langue française, Paris, Leipzig: Franck.
- Passy, P. (1891): Etudes sur les changements phonétiques, Paris: Didot.
- Richter, E. (1933): "Einheitlichkeit der Hervorhebungsabsicht", Actes Congr. Ling. 2.
- Romportl, M. (1973): "Zum Problem der Fragemelodie", Studies in Phonetics, Prague: Academia (147-164).

THE SOCIAL ORIGINS OF SOUND CHANGE

William Labov, University of Pennsylvania, Philadelphia, PA, USA

The past century of phonetic research has illuminated our understanding of the production of sounds, the properties of the acoustic signal, and to a certain extent, the perception of speech sounds.¹ Studies of the linguistic organization of these sounds have clarified our understanding of their distribution and diversification, the end results of the process of sound change. But the search for the originating causes of sound change itself remains one of the most recalcitrant problems of phonetic science. Bloomfield's position on this question is still the most judicious:

Although many sound changes shorten linguistic forms, simplify the phonetic system, or in some other way lessen the labor of utterance, yet no student has succeeded in establishing a correlation between sound change and any antecedent phenomenon: the causes of sound change are unknown. (1933:386)

In spite of Bloomfield's warning, linguists have continued to put forward simplistic theories that would attempt to explain sound change by a single formal principle, such as the simplification of rules, maximization of transparency, etc. But at the 2nd Congress of Nordic and General Linguistics, King rejected his own earlier reliance on simplification (1975), and recognized the point made 50 years earlier by Meillet (1921), Saussure (1922) and Bloomfield (1933): that the sporadic nature of sound change rules out the possibility of explanation through any permanent factor in the phonetic processing system. Explanations of the fluctuating course of sound change are not likely to carry much weight unless they take into consideration the parallel fluctuations in the structure of the society in which language is used.

The approach to the explanation of linguistic change outlined by Weinreich, Labov and Herzog (1968) divides the problem into five distinct areas: locating universal constraints, determining the mechanism of change, measuring the effects of structural embedding, estimating social evaluation, and finally, searching for causes of the actuation of sound changes. The quantitative study

 (1) The results reported in this paper are based on research supported by the National Science Foundation from 1973 to 1978. A more complete report is available in Labov et al., Social Determinants of Sound Change (1978).

of sound change in progress by Labov, Yaeger and Steiner (1972) located three universal constraints on vowel shifting, a line of investigation originally foreseen by Sweet (1888), and expanded the view of functional embedding in phonological space outlined by Martinet (1955). Our current studies of sound change in progress in Philadelphia have developed further techniques for the measurement and analysis of vowel shifts, with the end in view of attacking the actuation problem itself. We have approached the question of why sound changes take place at a particular time by searching for the social location of the innovators: asking which speakers are in fact responsible for the continued innovation of sound changes, and how their influence spreads to affect the entire speech community.

It is often assumed that sound change is no longer active in modern urban societies, and that local dialects are converging under the effect of mass media that disseminate the standard language. The results of sociolinguistic studies carried out since 1961 show that this is not the case: on the contrary, new sound changes are emerging and old ones proceeding to completion at a rapid rate in all of the speech communities that have been studied intensively. Evidence for sound changes in progress has been found in New York, Detroit, Buffalo, Chicago (Labov, Yaeger and Steiner 1972), Norwich (Trudgill 1972), Panama City (Cedergren 1973), Buenos Aires (Wolf and Jiménez 1978) and Paris (Lennig 1978). This evidence is provided by distributions across age levels (change in apparent time), and by comparison with earlier phonetic reports (change in real time), following the model of Gauchat 1904 and Hermann 1930.

Whenever these changes in progress have been correlated with distribution across social classes, a pattern has appeared that is completely at variance with earlier theories about the causes of sound change. If one looks to the principle of least effort as an explanation, or to discontinuities of communication within urban societies with accompanying isolation from the prestige models, then it would follow that sound change arises in the lowest social classes. Arguments for the naturalness of vernaculars and the marked character of prestige dialects would also look to the lowest social class as the originating site of sound change (Kroch 1978). If the theorist focuses on the laws of imitation (Tarde

1873) and the borrowing of prestige forms from centers of higher prestige, then it would follow that new sound changes will be the most advanced in the highest social classes. Neither of these cases has appeared in the internal changes studied in urban societies. It is true that older sound changes, like stable sociolinguistic variables, are often aligned with the socioeconomic hierarchy, so that the lowest social class uses the stigmatized variant most often, and the highest social class least often. But new sound changes in progress are associated with a curvilinear pattern of social distribution, where the innovating groups are located centrally in that hierarchy: the upper working class, for example, or the lower middle class.

Thus in New York City, lower middle class groups were the most advanced in the raising of long open o in lost, law, etc (Labov 1966, 1972). The same pattern was found in the backing of (ay) and the fronting of (aw) in that city. In Norwich, Trudgill found that the backing of short e before /l/ in belt, help, etc., showed a rapid development among younger speakers, and was most advanced in the upper working class (1972). In Panama City, Cedergren found that one of five sociolinguistic variables studied showed an age distribution characteristic of sound change in progress: the lenition of (ch) in cerca, muchacha, etc. This sound change showed a strong peak in the centrally located Classes II and III that Cedergren had established in Panama City (1973).

Our project on linguistic change and variation selected Philadelphia as a site for the further study of this problem, since it appeared that almost all of the Philadelphia vowels were in motion, and all of the basic patterns of chain shifting found in English and French dialects could also be located in Philadelphia. The main data base for the Philadelphia investigation is a series of long-term neighborhood studies in working class, middle class and upper class areas, involving repeated interviews and participant observation of the speech community. To this is added a geographically random survey of telephone users employing short, relatively formal interviews. The convergence of the findings from these two data bases, which show opposing strengths and sources of error, provides strong support for the general findings, though only data from the neighborhood studies will be presented here.

The measurement of vowel nuclei was carried out by a frequen-

cy analysis using a real-time spectrum analyzer (SD 301C), followed by linear predictive coding of the frequency domain (Markel and Gray 1976, Makhoul 1975) to derive more exact estimates of the central tendencies of F1, F2, F3 and F \emptyset . Complete vowel analyses of spontaneous speech were carried out for 97 subjects in the neighborhood studies and 60 subjects in the telephone survey, with 150-200 vowels measured for each subject. The mean values for each subject were then submitted to three normalization programs: a log mean model developed by Nearey (1977), the vocal tract scaling of Nordström and Lindblom (1975), and a three parameter method developed by Sankoff, Shorrock and McKay (1974).

Stepwise regression was carried out on the unnormalized and normalized series, deriving equations that predicted mean F1 and F2 positions from age, sex, social class, social mobility, ethnicity, neighborhood, communication patterns and the influence of other languages. The regression program enters into the equation the independent variable that has the highest partial correlation with the mean formant values, and with each successive term re-examines all previous terms as if they were the last to be added to the equation: if their effect falls below a given level of significance, they are removed (Draper and Smith 1966, Efroymson 1960). Thus the relative order in which variables are presented to the program is immaterial.

We then searched for the method of normalization that showed the maximum clustering to eliminate the effects of differences in vocal tract length, and the minimum tendency to eliminate variation known to be present in the data by independent means. Uniform scaling based on the geometric or log mean (Nearey 1977) was selected by these criteria and will be used as the basis for the discussions to follow.

Figure 1 shows the mean positions of the Philadelphia vowels of 93 speakers in the neighborhood series. It also shows vectors representing the significant age coefficients of the regression equations. The age coefficients are multiplied by the chronological age of the subject, e.g.

$$F2(aw) = 2086 - 5.39 \cdot \text{Age}^{[t=6.0]} \dots$$

where the numbers may be read as F2 values in Hz. Thus the first and most significant coefficient shown above predicts that the difference in mean F2 positions for two speakers 50 and 25 years old

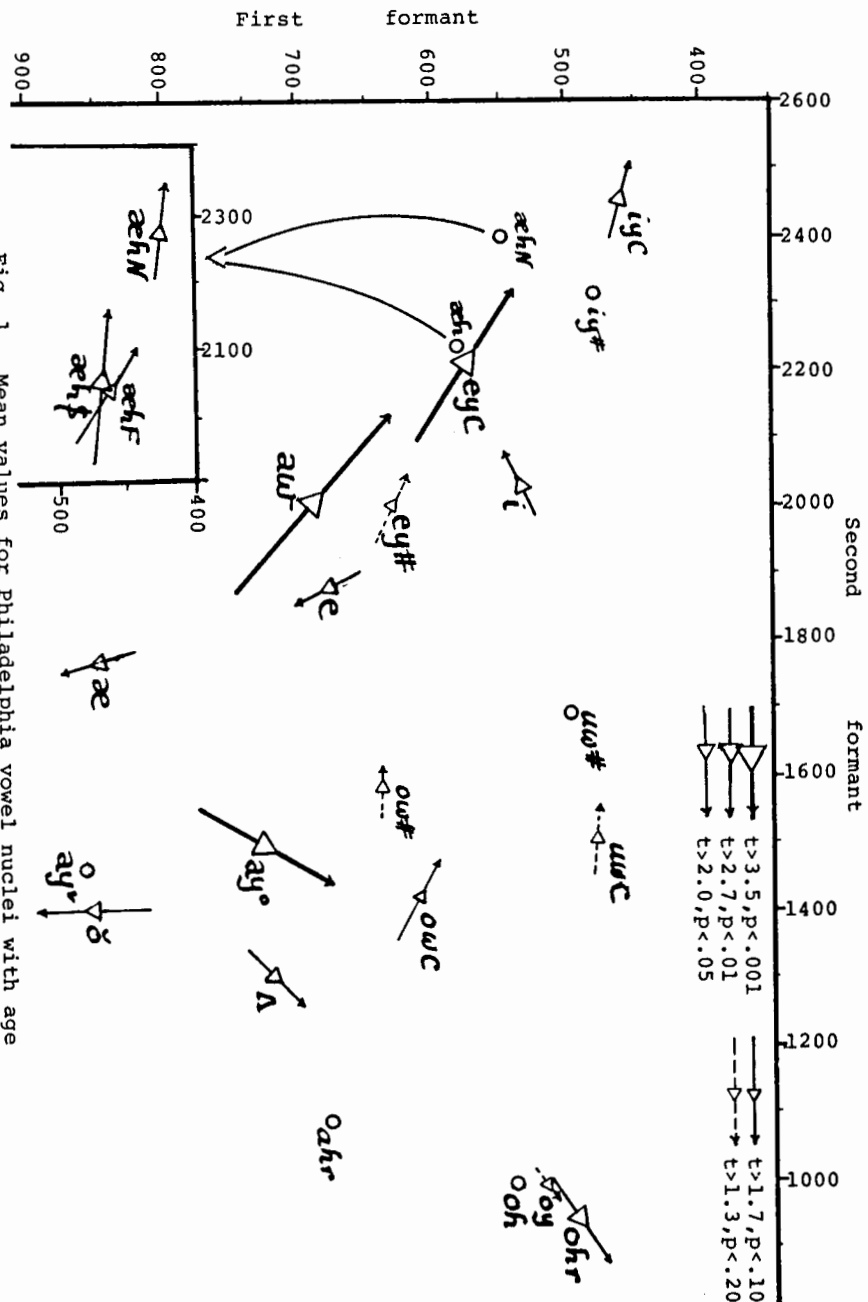


Fig. 1. Mean values for Philadelphia vowel nuclei with age coefficients shown as vectors projected 25 years ahead and 25 years behind the mean: data for 93 neighborhood speakers.

will be (25)(5.39) Hz: that is, the younger speakers will have a mean F2 135 Hz greater than the older. The vectors on Figure 1 represent the result of projecting the sound change 25 years ahead of the mean value and 25 years behind it. The significance of the effect is shown by the size of the triangles and the heaviness of the vector lines.

These age vectors fit in with evidence derived from earlier records and synchronic characteristics of the current data that allow us to set up five strata of sound change in Philadelphia:

- recently completed changes: e.g., the raising of /ahr/ in car, part, etc.
- changes nearing completion: e.g., the raising and fronting of (æh) in man, hand, etc.
- middle range changes: the fronting of (uw) and (ow) in too, moved, go and code (but not before liquids).
- new and vigorous changes, not reported in earlier records: the raising and fronting of (aw) in house, down, etc., from [æ^u] to [e^o]; the raising and backing of (ay^o) before voiceless consonants in fight, like, etc., from [a^l] to [æ^l]; the raising of (eyC) in the checked syllables of made, lake, etc., from [e^l] to [æ^l].
- incipient changes, e.g., the lowering of the short vowels /i/, /e/ and /æ/.

Conclusions from the earlier studies would lead us to associate a curvilinear social pattern with (d) the new and vigorous changes represented by the long, heavy vectors in Figure 1. Further terms in the regression equation show that this is the case. Extending the equation for F2 of (aw) to the three next most significant coefficients, we have [SEC = 'socio-economic class']:

$$F2(aw) = 2086 - 5.39 \cdot \text{Age} + 126 \cdot \text{Female}^{[t=3.5]} + 261 \cdot \text{SEC } 9^{[t=3.1]} - 253 \cdot \text{SEC } 13-15^{[t=2.5]}$$

In this socio-economic class scale, a 16-point index based on education, occupation and residence value, SEC 9 is generally considered the highest section of the working class, and 13-15 the upper middle class. Fig. 2 shows the coefficients for F1 and F2 projected as a single index on the front diagonal for all SEC, forming a smooth curvilinear pattern around SEC 9. Non-significant points are consistent with the main effects shown above.

Figure 3 shows the class distribution for the projection of checked (eyC). This is a broader curvilinear pattern with a significant peak in the middle working class group SEC 7, and two

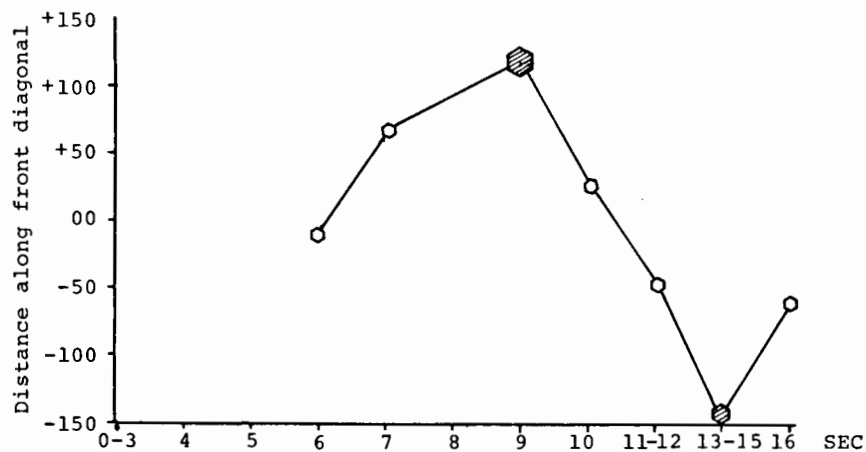


Fig 2. Projection along front diagonal of regression coefficients for F1 and F2 of (aw) for all socio-economic classes compared to SEC 0-3: data from 93 speakers in Philadelphia neighborhood study.

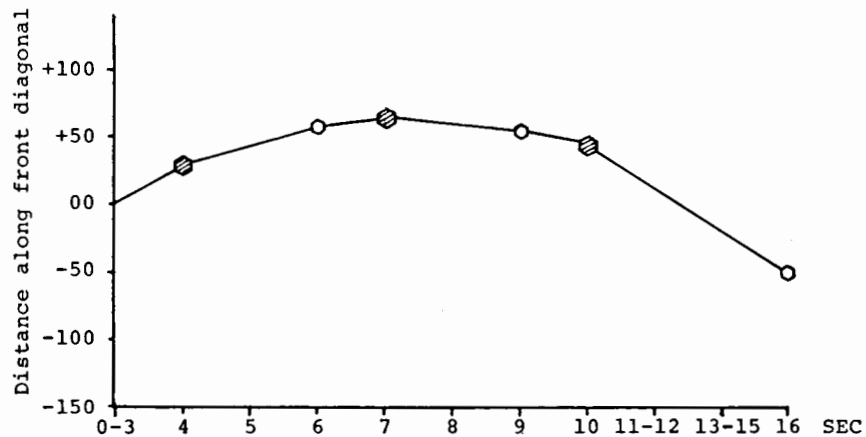


Fig 3. Projection along front diagonal of regression coefficients for F1 and F2 of (eyC) for all socio-economic classes compared to SEC 0-3: data from 93 speakers in Philadelphia neighborhood study.

- ▨ $p < .01$
- ◐ $p < .05$
- $p < .10$

other points significantly higher than the reference level of SEC 0-3, located symmetrically above and below SEC 7. Again, the less significant points form a smooth curvilinear pattern.

The third new and vigorous change, the raising of (ay⁰), shows no significant class distribution. It is worth noting that this is also the only change where men are in the lead: as in most previous studies of vowel change Philadelphia women are about one generation ahead of men in the early stages of change, except in the case of (ay⁰). Whatever the explanation for this connection between sex and SEC patterns, the Philadelphia results agree with impressionistic studies in showing no cases where the lowest or highest social classes appear as innovators in systematic change.

Given the powerful tendency for systematic sound changes to arise in interior social groups, we must ask how this observation bears on the causes and motivations of sound change. Instead of pursuing speculations on the psychological traits of these upper working class innovators, it will be more fruitful to probe more deeply into their social roles and relations to others in the community. The further investigation of the problem carried out by our research group is based on the evidence of communication networks which cannot be presented in this brief report. In general, it can be said that the speakers who are most advanced in these sound changes are those with the highest status in their local community: more specifically, they are persons with the largest number of local contacts within the neighborhood, yet at the same time with the highest proportion of their acquaintances outside the neighborhood. A portrait is beginning to emerge of the individuals with the highest local prestige who are responsive to the broader, almost metropolitan prestige that has become associated with the sound changes in question. It is plain that we are dealing with the emblematic function of phonetic differentiation: the identification of a particular way of speaking with the norms of a particular local community.

Through the further study of the role of new ethnic groups entering the community, and the communication patterns that connect local neighborhoods, we hope to delineate more closely the social pressures that are responsible for the dissemination and further advance of sound change, and thus isolate the driving force behind the continued diversification of linguistic forms.

References

- Bloomfield, L. (1933): Language, New York: Henry Holt.
- Cedergren, H.J. (1973): The interplay of social and linguistic factors in Panama. Unpublished Cornell U. dissertation.
- Draper, N.R. and H. Smith (1966): Applied regression analysis, New York: Wiley.
- Efroymson, M.A. (1966): "Multiple regression analysis", in Mathematical methods for digital computers, A. Ralston and H. S. Wilf (eds.), 191-203, New York: Wiley.
- Gauchat, Louis (1905): "L'unité phonétique dans le patois d'une commune", In Festschrift Heinreich Morf, 175-232, Halle: Max Niemeyer.
- Hermann, E. (1929): "Lautveränderungen in der Individualsprache einer Mundart", Nachrichten der Gesellsch. der Wissenschaften zu Göttingen, Phil.-his. Kl., 11, 195-214.
- King, R. (1975): "Integrating linguistic change", The Nordic Languages and Modern Linguistics, K.-H. Dahlstedt (ed.), 47-69, Stockholm: Almqvist & Wiksell.
- Kroch, A. (1978): "Toward a theory of social dialect variation", Language in society 7, 17-36.
- Labov, W. (1966): The social stratification of English in New York City, Washington: Center for Applied Linguistics.
- Labov, W. (1972): Sociolinguistic patterns, Philadelphia: University of Pennsylvania Press.
- Labov, W., M. Yaeger and R. Steiner (1972): A quantitative study of sound change in progress, Philadelphia: U.S. Regional Survey.
- Labov, W., A. Bower, D. Hindle, E. Dayton, M. Lennig and D. Schiffrin (1978): Social determinants of sound change, Philadelphia: U.S. Regional Survey.
- Lennig, M. (1978): Acoustic measurement of linguistic change: the modern Paris vowel system. Unpublished University of Pennsylvania dissertation.
- Makhoul, J. (1975): "Spectral linear prediction: properties and applications", IEEE Transactions on Acoustics, Speech and Signal Processing Vol. ASSP-23, No. 3.
- Markel, J. and A. H. Gray, Jr. (1976): Linear prediction of speech, Cambridge, MA: Bolt, Beranek and Newman.
- Martinet, A. (1955): Economie des changements phonétiques, Berne: Francke.
- Meillet, A. (1921): Linguistique historique et linguistique générale, Paris: La société linguistique de Paris.
- Nearey, T. (1977): Phonetic feature systems for vowels. Unpublished University of Connecticut dissertation.
- Nordström, P.-E. and B. Lindblom (1975): "A normalization procedure for vowel formant data", Paper 212 at the 8th Int. Cong. of Phonetic Sciences, Leeds.
- Sankoff, D., R. W. Shorrock and W. McKay (1974): "Normalization of formant space through the least squares affine transformation", Unpublished program and documentation.
- Saussure, F. de (1922): Cours de linguistique générale, 2nd ed., Paris.
- Sweet, H. (1888): A history of English sounds. Oxford: Clarendon Press.
- Tarde, G. (1873): Les lois d'imitation.
- Trudgill, P. J. (1972): The social differentiation of English in Norwich, Cambridge: University of Cambridge Press.
- Weinreich, U., W. Labov and M. Herzog (1968): "Empirical foundations for a theory of language change", in Directions for historical linguistics, W. Lehmann and Y. Malkiel (eds.) 97-195, Austin, Tex.: University of Texas Press.
- Wolf, C. and E. Jiménez (1978): "A sound change in progress: the devoicing of Buenos Aires /ʒ/ into /ʒ̥/", unpublished paper.

SOCIAL STRUCTURE AND PHONEMIC MODIFICATION

Bertil Malmberg, Dep. of General Linguistics, Östervångsvägen 42,
223 65 Lund, Sweden

When I made my first efforts to apply principles of Prague phonology in an analysis of the French and Italian vocalic systems (Acta Linguistica II, 1940-41, 232-246; III, 1942-43, 34-56), I soon arrived at the conclusion that this was not possible if I assumed that two units of expression (in my case two "vowels"), in a given position, and throughout the vocabulary, were either variants (allophones) or invariants (phonemes). It turned out that certain units were definitely phonological in some words, mere variants (free allophones) in others. The fact that /e/ - /ɛ/ in final position is definitely distinctive in pairs like dé-dais, fée-fait does not exclude their use as free variants in e.g. quai, gai, (je) sais. As far as the two a:s (/a/ and /ɑ/) are concerned, they have their full phonological value only in relatively few words (lâ-las). Even Parisians (only the language of the capital is referred to here) who agree on the existence of the opposition often do not agree on the distribution of the units in the vocabulary. I had also mentioned in my early study the critical oppositions /ø/ - /œ/ and, though better maintained, /o/ - /ɔ/ (both in closed syllable). I also mentioned the quantitative opposition, still retained by a few speakers, between /ɛ/ (mettre) and /ɛ:/ (maître). The cases of merger were far too frequent to be dismissed as mere phonemic word variants (Jones). I had drawn from these findings the conclusion that it was reasonable to look upon the French vocalism as containing two phonological systems, one richer and another poorer, or in my terms, a maximum system and a minimum system, one of them applied by certain speakers and in certain types of words, the other applied by others and in other words. I never concluded that some speakers used just one, others the other of those two systems in their entirety. I still do not know if there are native Parisians who make full use of the maximum system in any possible position and other native speakers who content themselves with the minimum one throughout the vocabulary. What is certain is, however, that the latter case seems to be normal in the pronunciation of numerous immigrants from the provinces and particularly in southerners and French immigrants from North Africa.

There seems to be no doubt that the choice between a more complex and a more reduced system is determined by non-linguistic (social, cultural, and in the case of immigrants from other French-speaking areas, regional) factors. The maximum system represents the complete set of oppositions permissible according to the paradigm and all the syntagmatic distinctions admitted by the distributional laws - the minimum system the smallest number of distinctive units without which the message does not function and the identification of the meaningful units ceases. (When putting it that way I do not take into consideration factors such as redundancy and context.) In other words, the difference between the two is one between what a speaker can and what he must do. The same interpretation seemed to me to be useful in the analysis of other complicated systems, i.e. the word accent problem in Scandinavia and (as demonstrated in the article quoted in Acta Linguistica III), the oppositions /e/ - /ɛ/ and /o/ - /ɔ/ in Italian, where in both cases it is a question of interference between dialects (or regional variants of the standard), whereas in French the situation can at least partly be interpreted as one between diachronically different systems (though both present at the same time and transformed into social or individual phenomena). Consequently, a state of language (introduced here as a translation of état de langue used in my French text,¹ a concept which goes back at least as far as Saussure's "Cours") may contain different strata, the most simplified of them pointing in the direction the evolution will take if no intervening factors prevent it. It is from this point of view that such an idea may be useful for a correct interpretation of diachronic, or evolutionary phonology. A language thus becomes a harmonious achronic system, or rather complex of systems, whereas a state of language is a linguistic situation described as valid for a chosen period of time or/and for a chosen spatial region or social stratum (all arbitrarily chosen).

The minimal system of French vowels represents a reduction in relation to the fuller one; in purely synchronic terms a system of inferior complexity. Diachronically it represents a loss of certain oppositions retained in the richer one. In all the cases under

(1) See my article in *Mélanges Straka I*, 1970, reprinted in Malmberg, "Linguistique générale et romane", Mouton, Paris 1973, 155-159.

discussion, the distinctions are phonetically subtle. This means that the oppositions based on the slightest differences of articulation and perception have been eliminated or, in most of our examples, reduced in their usage to a small number of words, forms, and contexts. This is typical of what happens in languages in reduction or destruction (in evolutionary phonology, in aphasia, etc.), and in reversed order in languages in construction (in the child, in the language learner, etc.). This is a consequence of Jakobson's law, implying that the complex system supposes the less complex ones, the subtle differences the rougher ones. We know that this law is valid in language learning and in language loss. It must necessarily be taken into consideration also in a study of linguistic change (phonological or other). We also know that the complete elimination of a language - under the pressure of another or owing to lack of motivation for its conservation - takes place according to the same hierarchic order. A situation such as the one reflected in the actual French vocalism is typical of a stage which precedes a generalized simplification. This does not mean that the simplification will necessarily take place. The choice of the speakers may be directed towards a retention of status quo, or even lead to a reestablishment of the more complex system (an example seems to be the opposition /e:/ - /ɛ:/ in the Swedish pronunciation of Stockholm).

My thesis is consequently that any state of language contains levels of different complexity from the maximum system maintained by strong linguistic norms, through degrees of increasing simplification down to the minimum system, and even beyond these to defect forms of language in the child, in aphasia, or in other disorders such as deafness, and in such foreigners and bilinguals as belong only partly to the socio-linguistic group in question. Any language system and, more generally, any semiotic system, is maintained thanks to a tradition respected by the members of society. Its basis is the prestige of norms regulating people's behaviour. The structural reduction of a system and its final elimination is the inverted function of the strength of the norms which guarantee its validity. Consequently, the existence of levels of varying structural complexity is due to the incapacity of the norm to maintain the complete system down to the lowest strata of society, in the more distant parts of the linguistic community, and under un-

favourable external conditions. Those are only aspects of the same phenomenon. In earlier studies and particularly with reference to Romance and Hispanic phonological evolution,² I have proposed to talk about simplification in the periphery. It follows from what has been said so far here that the concept of periphery is used with reference to two dimensions: spatial and social. The simplified or defect linguistic usage in the lowest social and cultural strata is peripheral in the same sense as the form of language in distant regions, far from normative centres. The concept of distance is consequently taken as meaning horizontal as well as vertical remoteness. With a slightly deviating use of the term it may even be extended to cover a weak (individual) mastery of the functional system.

If we look upon a state of language as a unity of systems of varying complexity, it will be necessary to introduce as a further variable the concept of choice. A Frenchman of today may choose one type of structure or another. His choice will be determined by his preference for one or another of existing norms (any linguistic usage being, of course, governed by some norm). He may make his choice unconsciously and in accordance with his social (cultural) background, or in a conscious intention to manifest his position as belonging to the upper ten, or as loyal to the social group where he comes from or to which - for personal or ideological reasons - he wants to belong. In such cases, his choice of pronunciation may function in exactly the same way as his choice of clothes or his social behaviour in general.

When, in my plenary report to the International Congress of Linguists in Bucharest (in 1967), I formulated the consciously and intentionally provocative thesis that language does not change and that what we call linguistic change is the speaker's choice of another language (taken here as a stable system of functions, and independent of any time factor), I thereby wanted to stress the importance of the choice factor in the evolution. I found it fruitful to see language as consisting of strata or levels, the choice between which is determined by social evaluations, even by changing

 (2) Summarized in *Orbis* XI, 1, 1962, 131-178 (reprinted in "Phonétique générale et romane", Mouton, Paris 1971, 301-342), and, as far as Spanish is concerned, in "La América hispano-hablante", Istmo, Madrid 1970.

modes. We have seen that modifications by choice may in principle take place in two directions: downwards, towards a simpler structure, and upwards, i.e. replacing a simpler structure by a more complex one. The danger of homonyms, often quoted as an important positive factor, and the absence of them as a negative one, has probably been exaggerated.

Interference (substratum, superstratum, adstratum) has often been quoted as an underlying factor in sound change. It supposes bilingualism. Bilingual areas and societies are given as examples of conditions under which the linguistic norm may be weakened and where system reductions a priori seem probable (well-known examples are the loss of voiced stops in the French spoken in Alsace and the loss of the phonemic word accent in the Swedish of Finland). Now an important question arises: are such peripheral simplifications to be explained through direct influence from the language which ignores the distinction, or are they simply due to a general weakening of the norms in a peripheral area? We know that voiced stops are relatively rare and that they come late in the child's linguistic development. We also know that phonological word tones belong to the subtle phonological distinctions, late in Swedish children and absent in cases of individual linguistic weakness. This question can hardly be answered. The effects are the same. When the change is just a phonological reduction, the interference theory is superfluous. Only when the new system contains new structural features and/or structural relations do we have any real reason to consider an interference theory. The introduction into Northern Gallo-romance of the phoneme /h/ as a consequence of the Frank colonisation (retained till today in some dialects, Normandy) is inexplicable without the foreign influence (and understandable in consideration of the socio-linguistic situation in the bilingual Frank kingdom).

It seems, on the other hand, quite normal if in a language in close contact with a quite different neighbouring one whose influence on the former is understandable (socially, culturally, politically, or simply through a quantitative dominance), we meet phenomena of phonetic realization of the phonological system which have to be explained through interference between different speaking habits. The examples are numerous (the lack of aspiration of Swedish-Finnish /p, t, k/; the pronunciation of the Spanish /j/-

phoneme as /d̃j/ in Paraguay; intonation and stress phenomena). These features do not belong to the phonological system strictly speaking (though they may play a part in communication on other levels than the strictly cognitive one). And they may come to play a role at later stages in the phonological evolution (an example later).

In my critical studies on Romance diachronic phonology, I have been very restrictive as far as interference theories are concerned. I have tried to prefer internal evolution and peripheral simplification as explanations, the latter socially determined.

The expansion of Castilian in medieval Spain which became a consequence of the reconquest ("reconquista") from the Arabs, as well as its continuation (from 1492) in America, implied numerous instances of structural simplification of the phonological system (loss of the medieval opposition between voiceless and voiced fricatives, voiced and fricative stops). This evolution was parallel with the social changes brought about by the political events. The medieval /ts/ was replaced by the interdental /θ/ in the centre but confused with /s/ in the South and in America ("seseo"). A widespread dialectal confusion of liquids is found in (regionally and/or socially) peripheral strata all over the Spanish speaking world. It results in a substitution of one for the other (mostly a generalization of l), or in a phonetically intermediate type. A map published by Alonso-Lida (Rev. Fil. Hisp. VII, 1945, 320) of the extension of the merger in Spain shows its marginal character. Other phenomena of simplification show a corresponding spatial and social extension on both continents. The Spanish of America reflects the differences of political, social, and spiritual structure in the colonial period. The replacement of implosive -s in Spanish through an undifferentiated h-like fricative has the same extension as other "vulgarisms", in Spain and in America. Though it is a mere manifestation of the s-phoneme, it may have secondary phonemic consequences (lengthening of vowels, change of vowel quality) and ought to be mentioned for this reason. A parallel evolution took place in medieval French and is still reflected in oppositions like Fr. patte - pâte.

Linguistic evolution would not be conceivable without the hierarchic differentiation of a state of language, without vari-

ability in the strength of norms, and without a choice (free within limits) made by the members of the different social strata. These are the essential factors in the socio-linguistic evolution.

In conclusion: diachrony interpreted as a substitution of one system for another (in any of the dimensions of language) through a socially determined choice between possibilities of varying complexity was the principle I wanted to submit for consideration to the Bucharest Congress of 1967. I did it by saying: language does not change; man changes languages.

THE REALITY OF SOUND CHANGE: A SOCIOLINGUISTIC INTERPRETATION

Fred C.C. Peng, International Christian University, 10-2, 3 Chome, Mitaka, Tokyo, 181, Japan

This paper attempts to summarize the latest findings of my research on sound change. It also contains criticisms of and comments on previous studies along this line. In the main, a new theory is proposed, suggesting that the process of sound change can be observed within one generation. Given this theory, four questions are asked, which become the focus of my argument in the course of discussion.

The Problem

Sound change has been an intriguing subject in general linguistics for almost two centuries. I wish to emphasize, however, that language as a code does not change by itself; people who employ the code change it. It is from this point of view that I shall address myself to the reality of sound change.

To begin with, let me identify the problem. Linguists have in the past been led to believe that it will take generations to produce certain changes and that the length of time that is needed to show such changes is too long, or to put it the other way around, that the ongoing progress of such changes is too subtle and slow to allow any direct observation.

Labov has recently challenged this traditional belief by advocating that change can indeed be directly observed. However, Labov's observation of sound change in Martha's Vineyard, involving a claim that sound change may be captured while in progress, takes in three generations (Peng 1976, 70), thereby yielding to the "myth" in the literature that changes occur across the boundaries of two or more generations.

This myth was repeated once more by Johnson recently (1976) who claims that "The time span considered can be across several centuries or as few as two demographic generations" (1976, 165). He thus concludes that "Specifying the terms 'fast' and 'slow', we have given some support to the claim that change begins slowly and accelerates in succeeding generations, and we have given evidence that change advances more rapidly in urban than in rural communities" (1976, 171).

In view of this (unfortunate) development, several questions need to be raised here, so as to eradicate once and for all the

myth that seems to persist in the literature. For the sake of convenience, these questions are asked below in the order in which I shall discuss them in this paper:

- Q1. Is it linguistically plausible to construct a theory of sound change that is based on the assumption that sound change takes place across the boundaries of two or more generations?
- Q2. Is it true that change begins slowly and accelerates in succeeding generations?
- Q3. Is it theoretically sound to generalize from one type of changes in one language to the same type of changes in other languages?
- Q4. Can linguists, historical linguists in particular, do themselves justice by ignoring nonlinguistic changes when they deal with linguistic change?

Previous Study on Sound Change within One Generation

Let me quickly review what I said in Peng (1976) concerning sound change within one generation. First I took Nomoto's 1950 study and 1971 study and came up with the result that each individual seems to continue developing his or her speech beyond 13 years of age, at an ever decreasing rate, until the age of 35 or thereabout. I added that "Such is the case in most of the phonetic parameters" (1976, 82). I then proceeded to ask a question: If changes can be directly observed to take place within one generation, what are the mechanisms of sound change that may be discerned from the study? Five mechanisms were then singled out: Age factor, Educational background, Phonetic parameter (i.e., the choice of speech sound), Oscillation (for what Weinreich called retrograde), and Life expectancy. Each mechanism was elaborated on the basis of supporting data (1976, 83-90).

Second, I took Jespersen's metaphor and compared it with my alternative schematic representation of language change, which may be recapitulated as follows (1976, 91):

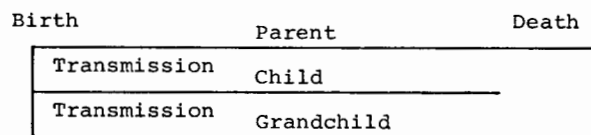


Figure 1. Illustration of Jespersen's Metaphor

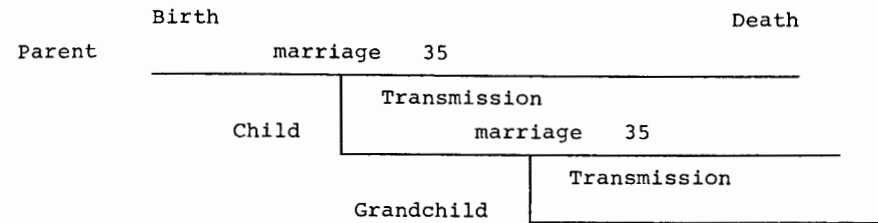


Figure 2
Alternative Schematic Representation
of Language Change

This alternative theory suggests that the child can learn his language perfectly; that in spite of his perfect learning, the language in question still changes because the model the child learns his language from had changed considerably before the child was born; and that the child's model had, in turn, learned from quite a different model, just as the child will serve as quite a different model to his own child. In this way, I concluded that a sound change, be it abrupt or not phonetically, can only be gradual in terms of behavior within each individual, with smooth (i.e., perfect) transmission from generation to generation (1976, 92).

To illustrate this point, let me make a distinction between changes in language behavior and changes in linguistic code. This distinction is important because the accumulation of changes in language behavior results in changes in linguistic code, and changes in language behavior are directly observable; on the other hand, changes in linguistic code may or may not be so observed if one's aim is to determine the end points, rather than the ongoing processes, of the period of operation of a sound change. An exemplification of this distinction is in order here.

In his criticism of the 'gradual view', Wang cites an interesting case as follows: "so, for a word like acclimate in which the pronunciation changes from [əkliájmt̩], the only pronunciation found in some older dictionaries, to [əkliúmejt̩], where all three vowels are different (in addition to the difference in accent pattern), it is surely unrealistic to suppose that there was a gradual and proportionate shift along all four phonetic dimensions" (1969, 14).

Note that while the change from the first to the second pronunciation may be abrupt along all four phonetic dimensions or even one phonetic dimension, it is notwithstanding a change in the

system of the linguistic code. Thus, the abruptness is immaterial here, because any native speaker of English can switch instantly from one pronunciation to the other with little difficulty.

By contrast, however, the change in language behavior from the first to the second pronunciation must be gradual. This aspect of gradualness can be directly observed and measured as part and parcel of language behavior, among various groups of people with varying social backgrounds.

From the above review it must now follow that if changes in language behavior can be systematically described, there is no need to wait for the result (i.e., the end point) to show up in the code itself. We must come to grips with the ongoing process of changes in language behavior that underlie the net result of changes (i.e. end points) in the linguistic code.

Discussion

With the conception of sound change presented above in mind, let me now return to the questions originally asked. First, I must mention that it is rather unfortunate that Johnson repeats the traditional view that sound change must take place across generation boundaries.

Empirical evidence is presented in Peng (1976 and n.d.) that sound change takes place not only within each individual but at an ever decreasing rate, that is, taken cross-sectionally, a person may change his linguistic system within his life span but gradually reduces his rate of change until the age of 35, even though changes may continue to take place after the age of 35 (but at a much reduced rate). In light of this finding, it is hard to believe that sound change must take place across generation boundaries.

Second, given the above finding that sound change takes place within each individual at an ever decreasing rate, I must now ask whether it is true that change begins slowly and accelerates in succeeding generations. Although the data presented by Johnson may seem suggestive of this tendency, a closer look at his data indicates otherwise (especially when they are compared with ours), simply because ours can account for changes within one generation, whereas Johnson's (which include several sources) contain materials from at least three generations, each having a different age bracket and being younger than the preceding generation. For instance, he uses Labov's material from Martha's Vineyard (aw) that covers three generations; namely, Oldest Generation, Middle Genera-

tion, and Youngest Generation. But note that the three generations correspond to age level 61 to 90, age level 31 to 60, and age level 30 and under, respectively (cf. Labov 1972, 22 and 279), and that there is no information about the changes that the younger age groups will exhibit when they reach the older bracket. Thus, when the numerical values (Johnson 1976, 168) of 0.06, 0.37, and 0.88 are compared, the differences do not represent the acceleration of change rate in three succeeding generations; rather, they indicate three static manifestations of one continuous change taken cross-sectionally. In order to get the dynamics of change, what Johnson should have done would be something like this: Wait for the people of the younger generation to reach the next age level (i.e., 30 years) and then compare their centralization with that of the older generation at the same age level. For instance, he should have got the numeric value of the Youngest Generation (under 30) when they reach the next age level (31-60) and compare it with the numeric value of the Middle Generation when they are still at the level of 31-60 and do likewise for the Middle Generation and the Oldest Generation. But nothing of this sort has been done. Consequently, he has no data whatsoever to support the claim of acceleration in the rate of change.

By contrast, our data from the area study show exactly this kind of dynamics pertaining to change. That is, the results of all age groups investigated in 1950 were compared, 21 years later, with those of similar age groups investigated in 1971. Thus, we have information not only on two comparable age groups, say, 35-44, one taken from the 1950 study and the other from the 1971 study, for comparison pertaining to change, but also on different age groups taken cross-sectionally for comparison pertaining to the rate of change. The data from the area study are then backed up almost one to one by our data from the panel study. Thus, in the case of sound change, we can comfortably conclude that all age groups have changed but that the rate of change goes down as the age goes up within each generation.

From the aforementioned it must follow that there is a certain degree of incongruity in Johnson's data. For instance, how can he be sure that the first generation (Oldest Generation) did not have a faster rate of change when they were younger and that the third generation (Youngest Generation) will not slow down when they grow older? In fact, his data support precisely what we have found if

his three generations are regarded cross-sectionally, which is to say that the rate of change will be reduced in all cases, e.g., Martha's Vineyard, as one goes from the Youngest Generation (0.88) through the Middle Generation (0.37) to the Oldest Generation (0.06). The fact that Johnson has no data for each individual within one generation (which, by contrast, we have in the panel study) regarding his or her changes suggests that he cannot be sure of the rate of change being faster in each succeeding generation, that is, accelerating in succeeding generations. To demonstrate this fact, let me resort to a schematic representation of language change.

Figure 3 depicts sound change within one's own life time (notably from 15 to 44 years of age) with a plotted extension (dotted line) beyond 44. I have also circled three places which correspond to Labov's three age levels utilized in Johnson's data. The result of these modifications in the schematic representation is recapitulated as follows:

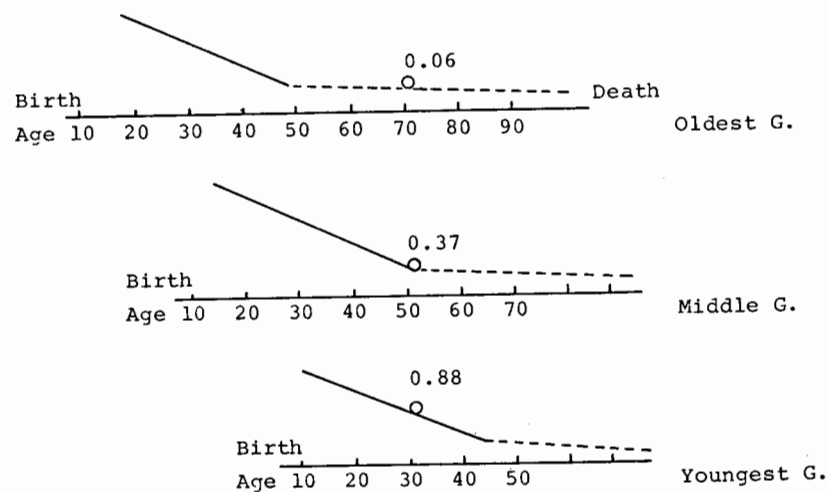


Figure 3
Schematic representation of sound
change and its rate

Observe now that this schematic representation shows that what Johnson has done is pick the three age brackets, one from each generation, with differing numeric values of vowel centralization (each of which falls in line with and can be explained by the rate of sound change therein). From my point of view, then, that the Oldest Generation has the lowest numeric value is not because, as Johnson has claimed, change begins slowly at first but because the age bracket (61-90) picked has, according to the schematic representation, already slowed down the rate of change; and likewise, that the Middle Generation and the Youngest Generation have successively increased their numeric values may also be explained by the fact that in the schematic representation they are younger in age and, therefore, stand higher in the rate of change. Consequently, it is not at all because change begins slowly and accelerates in succeeding generations, as claimed by Johnson who also thinks that he lends support to the claim of Wang and Cheng (in their discussion of lexical diffusion) and of Bailey that sound change follows an S-curve (Johnson 1976, 168). (By an S-curve is meant that sound change begins slowly and then increases rapidly [in some cases leaving residue].) In the light of my explanation above, it should be clear that none of the assertions made by Johnson and others is true.

At this point, I must add that certain sounds are more susceptible to changes than certain others (Peng 1976, 84 and 90). Given this view, which is supported by factual data from Japanese, the rate of sound change cannot be taken to mean that all language sounds (in a given language) progress in the same direction or at the same pace. Neither is it the case that the same type of sounds (in different languages) should have a fixed rate of change.

English may be cited as an example which shows marked ongoing changes in vowels rather than in consonants. This is, of course, historically true as well. However, another language, like Japanese, does not necessarily follow suit; my own study (Peng, 1976) clearly suggests that in Japanese consonants are much more susceptible to change. Thus, to answer Q3, I must say that whatever there is to discover regarding sound change in progress based, say, on English vowels, cannot and must not be generalized to apply to another language, unless there is a very good factual ground on which to build such a theoretical construct. I hope historical linguists have learned the lesson from the past, never to repeat the same mistake in the current exploration of sound change.

Conclusion

Let me now summarize by presenting three points, so as to bring the whole presentation to a close. Firstly, although linguists have been aware that when we speak of change it is people who change, and sound change is simply a manifestation (or symptom) of human change, not enough research is being done in, or attention paid to, the probe of what I have called the dynamics of change. This kind of study requires both cross-sectional and longitudinal investigations of fairly large samples in the same areas with the same method at an interval of hopefully 20 years. Since research of this nature is often painstaking and costly, historical linguists should turn to linguistic geographers and other social scientists for assistance in the provision of advice and materials; in spite of linguists like Kuryłowicz, who once renounced all support from linguistic geography and other social sciences for internal reconstruction (1964), it is through this kind of cross-fertilization that language scientists can hope to achieve the goal of dealing with the dynamics of change, among other things.

Secondly, I have also presented sufficient evidence to support my claim that if it is people who change, the change itself must take place within each individual to begin with, whose rate of change is affected by his or her own physical condition (age or maturation) as well as by the environment. Thus, as each individual increases his or her age, the rate of change decreases. Nobody knows, however, what will happen if life expectancy is extended beyond 100 years of age, to the rate of change.

Of course, life expectancy alone is not the influencing factor of human change; the environment counts heavily in this regard, the foremost influencing factor in the environment being human interaction. Note here that although the Japanese now live longer (perhaps longest?), 57% of the Japanese population is crowded on only 2% of the land, according to the latest report prepared by the Prime Minister's Office (The Japan Times, June 27, 1977). In this respect, then, Johnson is probably right in saying that "change proceeds more rapidly in urban than in rural areas" (1976, 165). I have reached a similar (albeit more substantial and elaborated) conclusion (Peng 1978).

Finally, I should mention that if human change is the key to sound change, more rigorous research is needed in such realms of

specialization as phonetics, neurolinguistics, sociolinguistics, and pedolinguistics to help determine the change and development in the total behaviors of humans as organisms.

References

- Johnson, Lawrence (1976): "A rate of change index for language", Language in Society 5, 165-172.
- Kuryłowicz, Jerzy (1964): "On the methods of internal reconstruction", in Proceedings of the Ninth International Congress of Linguists, 9-31, H.G. Lunt (ed.), The Hague: Mouton.
- Labov, William (1972): Sociolinguistic Patterns, Philadelphia: University of Pennsylvania Press.
- Nomoto, Kikuo et al. (1974): Chiiki Shakai no Gengo Seikatsu (Language behavior of a speech community), Report 52, Tokyo: National Language Research Institute.
- Peng, Fred C.C. (1976): "A new explanation of language change: The sociolinguistic approach", Forum Linguisticum 1, 67-94.
- Peng, Fred C.C. (1978): "Urbanization and language sciences: The Japanese case", in Language in Context, Fred C.C. Peng (ed.), Hiroshima: Bunka Hyoron Publishing Company.
- Peng, Fred C.C. (n.d.): "Sound change and language change: A sociolinguistic overview", special lecture delivered at the 1978 Annual Conference of the Linguistic Society of Japan (forthcoming in Language Sciences, 1978-9, vol. 1).

TEMPORAL RELATIONS WITHIN SPEECH UNITS

Summary of Moderator's Introduction

Ilse Lehiste, Department of Linguistics, Ohio State University

The title of the symposium leaves open the question of the type and size of the speech units. The contributors to the symposium have indeed chosen to address themselves to units of quite different types and sizes. Likewise, they have approached the problems connected with the temporal structure of speech units both from the perspective of speech production and from that of speech perception. The contributions include highly theoretical papers, papers presenting detailed results of experiments, and papers falling between these two poles. Some systematization appears to be in order. I would like to present herewith a framework within which I believe the issues can be profitably formulated for the discussions which I hope will follow.

The framework involves three dimensions. One of them concerns the relationship between timing control in production and the role of timing in perception. The second dimension deals with the direction of determination in the temporal organization of spoken language; specifically, with the question whether the timing of an utterance is determined by its syntax, or whether there exist rhythmic principles in production and perception that are at least partly independent of syntax. The third dimension follows directly from the previous two and relates to the type and size of speech units. What is the nature of those units, and are they to be established on the basis of a morphosyntactic analysis of the sentence, or on some kinds of independent phonetic criteria?

Clearly both production and perception are involved in oral communication by spoken language, and it would seem unnecessary to elaborate the point. However, I have had occasion to argue--against considerable weight of opinion--that durational differences in production, be they ever so significant statistically, cannot play a linguistically significant role if they are so small as to be below the perceptual threshold. It would be wise, I think, to remind oneself periodically of "the evident fact that we speak in order to be heard in order to be understood" (Jakobson et al. 1952). I hope, therefore, that in our discussion of temporal relations within speech units, models of production and models of perception will be related to each other.

The second and third questions concern the direction of determination: does phonology follow syntax, or are we dealing with interacting, but parallel hierarchies? Some researchers have developed programs for generating the temporal structure of a sentence on the basis of segments and syntactic structure, without paying any attention to rhythm. This is, I believe, due to a particular theoretical orientation. Generative phonology operates with segmental features; even suprasegmental features are attached to segments. And in a generative grammar, phonetic output is the last step in the generation of a sentence. An independent rhythm component simply has no place in the theory. For those scholars, then, the speech units are segments, phrases, clauses, and sentences. (And it is quite interesting to see them struggle with units not foreseen in the theory, like syllables and phonetic words.) Researchers who are not fully committed to this theoretical viewpoint operate with certain other units, such as speech measures or metric feet. Again, the reality of both kinds of units can be studied from the point of view of production as well as from that of perception.

Practically all the issues I have outlined are treated in the papers contributed to this symposium. Production is the main concern of the papers of Allen, Bannert, Klatt, and Öhman et al.; perception is the focus in the papers of Carlson et al., Donovan and Darwin, Fujisaki and Higuchi, Huggins, and Nootboom.

Among the papers dealing with production, Bannert considers the effect of sentence accent on the duration of VC sequences, employing a rather complex concept--vowel-to-sequence ratio $V/(V+C)$. The relationship between the VC-unit and its two parts represents a measure of the temporal structure of quantity of complementary length. Bannert shows that this unit is useful in describing the effect of the addition of sentence accent to quantity in Stockholm Swedish; it remains to be demonstrated whether the unit is as significant for perception as it is for production.

The paper by Klatt presents a detailed scheme for the synthesis by rule of segmental durations in English sentences. It is an almost pure example of the approach that starts from an abstract linguistic description and ends up as a sequence of segments, whose durations are conditioned by other segments and by syntactic con-

straints. Interestingly, a companion paper by Carlson et al. testing the output of Klatt's synthesis algorithm arrives at the conclusion that certain aspects of the durational pattern have greater perceptual importance than others. Vowel duration is more important than consonant duration; the durations between stressed vowel onsets seem to constitute a particularly important aspect of sentence structure.

The papers by Öhman et al. and by Allen concern themselves with production models in general. Öhman's et al. paper argues for a gesture theory of speech production. Their examples deal primarily with the assignment of fundamental frequency and are thus somewhat outside of the current topic. Allen's paper draws a useful distinction between descriptive models and theoretical models of speech timing, and makes the intriguing prediction that theoretical models may be about to undergo substantial modification, primarily due to the emergence of an "action theory" of speech production. According to that theory, neural activity is hierarchically organized into successively higher levels of coordination, until the highest level of all can only be described in terms of the overall goal of the action.

Among the papers devoted primarily to perception, Nootboom presents a decision strategy for the disambiguation of vowel length in Dutch. The strategy is complex, but listeners are fully capable of applying it in ongoing perception. Fujisaki and Higuchi present an analysis of the temporal organization of segmental features in Japanese disyllables consisting only of vowels, and find that although the onsets of the transition for the second vowel are distributed over a relatively wide range, a perceptual analysis of the onset of the second vowel shows relatively little temporal variation. It thus seems that the apparent diversity of the onset of transition in various disyllables is introduced to maintain the uniformity of perceived duration of segments. Fujisaki and Higuchi consider their results supportive of a model in which the motor commands and the articulatory/acoustic realizations of successive segments are programmed in such a way that the perceptual onsets of successive segments are isochronous.

The last two papers are likewise concerned with speech rhythm. Huggins finds that a correct rhythmic pattern, which is basically isochronous, enhances intelligibility, while a badly distorted

timing pattern impairs it seriously, even though all phonemes are identifiable. Donovan and Darwin deal with the perceived rhythm of speech, and give special consideration to the problem of isochrony. Their paper tests, among others, a hypothesis that I had formulated in 1973 and discussed in more detail in 1977. My observation was that listeners tend to hear utterances as more isochronous than they really are, and that listeners perform better in perceiving actual durational differences in non-speech as compared to speech. I concluded from this that isochrony is largely a perceptual phenomenon. Donovan and Darwin have confirmed these results. They make two points in addition: first, that isochrony is a perceptual phenomenon which is not independent of intonation, and second, that it is a perceptual phenomenon confined to language--reflecting underlying processes in speech production. Donovan and Darwin question the value of seeking direct links between syntax and segmental durations rather than indirect ones by way of an overall rhythmic structure.

I should like to propose a few direct questions for starting the discussion. What is the relationship between rhythm and syntax? How should rhythm be integrated into models of speech production and perception? What are the physiological constraints within which the production and perception of temporal structure must take place? What, indeed, is the nature of the temporal relations within speech units?

References

- Jakobson, R., C.G.M. Fant, and M. Halle (1952): Preliminaries to Speech Analysis, Cambridge, Massachusetts: MIT Press, tenth printing, 1972.
- Lehiste, I. (1973): "Rhythmic units and syntactic units in production and perception", JASA 54, 1228-1234.
- Lehiste, I. (1977): "Isochrony reconsidered", JPh 5, 253-263.

FORMAL AND STATISTICAL MODELS OF SPEECH TIMING: PAST, PRESENT,
AND FUTURE

George D. Allen, Dental Research Center, University of North
Carolina, Chapel Hill, N.C. 27514, USA

Let us begin this paper, whose goal is to review the kinds of models that have been developed in studies of timing in speech production and to suggest some possible directions for further research, by addressing briefly the general nature of models. Although the usual sense of the word "model" is that of an analogy, there is much room for differences in usage. On the one hand, we can have a "descriptive model", which models a set of observations, or data; a full-blown theory, on the other hand, models a complex and usually interacting set of constructs. Intermediate between these two extremes lies the single hypothesis, which is a projection of a theory onto a subspace of smaller dimensionality (often a single dimension) and which is "tested" by comparing it with a set of data. There are no theoretical boundaries between these three types of models, and most studies which model some aspect of speech timing contain elements of two (but seldom all three) of these categories.

Besides differing in the complexity of the structures which they reflect, models also differ in the intended accuracy with which they reflect those structures. Some models, for example, are intended primarily as conceptual guides, with only a loose fit between them and any existing data. Such models motivate the design of further studies and the analysis of data gathered by them, with the usually explicit goal of validation and refinement of the original model. The "chain" and "comb" models suggested by Bernstein (1967) were of this sort and have served as the basis for many recent studies of speech timing control. Other models are tailored closely to data or some other real world phenomena, their intent being more to parameterize the data (for example to permit comparisons of these parameters between different groups of speakers) than to explain the process whereby the data are generated. Klatt's (1975) study is an example of this kind of data-matching model. As with the descriptive vs theoretical distinction mentioned above, these extremes also allow much room for differences among models: about the only commonality among models in their "goodness-of-fit" to the data is that no model fits as well as its proponent would like.

A third difference among models is what I have chosen to term "formal" vs "statistical", though here again there is no true boundary between them. This contrast is exemplified by the difference between "regression" and "correlation" in statistics, the first being used to describe the form of the relationship between two measures (e.g., if A is 10 cm taller than B, then A may be expected to weigh about 5 kg more), the second an estimate of the strength of that relationship (e.g., A will weigh 5 ± 1.4 kg more, 95 percent of the time). Lindblom and Rapp (1973) have thus in this sense developed a formal model of segment duration, whereas Kozhevnikov and Chistovich (1965) carried out the first of many statistical studies seeking significant negative correlations between the durations of successive segments as evidence of temporal compensation within production units.

Let us now review some past and present models of speech timing and its control in terms of these different general features. This review unfortunately cannot begin to cover the wealth of studies that now exist in this area. It would be useful, for example, to try to relate models of production to models of perception. Instead I shall restrict attention here to just a representative sample of models of timing in speech production and hope that perhaps the symposium itself will bring about the more complete discussion this topic deserves.

One major class of models has concerned the durations of segments, the earliest studies dealing with vowel duration in English (House and Fairbanks, 1953; Peterson and Lehiste, 1960; Kim, 1966). Although all of these were primarily descriptive in their origin, Kim's was the most theoretical in its intent. By explicitly labeling the branches on his tree with fixed durational values to be attributed to plus- vs minus-tense vocalic nuclei or plus- vs minus-voicing of the arresting consonant, he cut some of the ties his model had to the data which generated it and aligned it as well as he could with the constructs of distinctive feature theory.

More recent models of segment duration are those of Lindblom and Rapp (1973) and Klatt (1975), mentioned earlier. Both of these are descriptive models, though the data they describe, and thus their derivative models, are different. Lindblom and Rapp used phonologically restricted nonsense material and described variations in a segment's duration as a function of the number of seg-

ments, syllables, and words following it in the phrase. Klatt, on the other hand, used a meaningful paragraph, sacrificing control over word- and phrase-length comparisons while retaining contrasts in local segmental and prosodic context and adding syntactic contrasts. Interestingly, both of these studies describe segment variation as a contextually conditioned reduction in duration from a longest "base" form; several other related studies (e.g., Nootboom, 1972; Umeda and Coker, 1975) have done the same, and Keating and Kubaska (1978) have suggested a role for this process in speech development.

Although these carefully constructed models are in substantial agreement as to the major dimensions required for describing the durations of segments in the phonologically restricted speech samples from which they were derived, other investigators have suggested that they do not model "real" speech. Umeda and Coker (1975), for example, present an alternative model, based again on measured segment durations but from corpora that are less constrained by laboratory conditions than, say, Lindblom and Rapp's (1973) or Nootboom's (1972) data. Their data, and therefore their model, show the same local contextual effects as the others' (e.g., neighboring segment and syllable types, degree of stress, syntactic word classes), but the longer term effects (number of syllables remaining in the word, and words remaining in the phrase) are absent. This difference shows clearly one of the principle hazards associated with models derived from data: an apparently important component or dimension of the model may turn out to be an artifact of the observational situation. In this particular case the issue remains open.

There are many other studies of segment duration that deserve recognition here, and much more that might be said concerning those studies which have been mentioned. Because of space limitations, however, let us move on to a second major class of speech timing models, those which have dealt with the control of the articulatory time program. Aside from the oversimplified but heuristically useful "isochronic" model of English stress (cf., e.g., Pike, 1945), the first model of speech timing control appeared in Kozhevnikov and Chistovich (1965). As noted earlier theirs tried, via statistical techniques, to identify temporal compensation within production units, their underlying goal being to validate either the "chain" or "comb" model proposed by Bernstein (1967). Because of

procedural artifacts inherent to their method, however, they recognized that they could not decide the issues from their data, and so they abandoned the temporal domain in favor of the articulatory. Some later investigators (e.g. Lehiste, 1972; Wright, 1974) were not so cautious and claimed evidence for temporal compensation in spite of warning by Kozhevnikov and Chistovich (1965) and Ohala (1970) that variations in speech rate and measurement error could mask any true effects. Allen (1973, 1974), on the other hand, tried to circumvent the methodological problem by proposing a statistical model which used a statistic that was insensitive to rate variations and by including an explicit estimate of measurement error. In agreement with Ohala (1970, 1975) he found no evidence for temporal compensation within the freely spoken phrase, thus supporting the "comb" model (though only weakly, since a statistically negative result can never be strong evidence for any hypothesis).

In addition to examining the relative validity of the "chain" and "comb" models, Allen's model had the additional advantage of yielding a measure of the speaker's timing control accuracy. In one study (Cooper and Allen, 1977) this model was partially validated using speakers whose timing control was known to be poor, and in another (Tingley and Allen, 1975) the developing ability of children's speech timing control was charted. As a result of these limited successes, Allen (1978) suggested that the methodological limitations inherent in earlier statistical approaches to the study of speech timing control may yet be overcome.

Although Kozhevnikov and Chistovich (1965) and most other investigators were seeking to discover units of speech production, Allen (1973) was equally interested in determining the nature of the mechanism for speech timing control. Following Huggins (1972), Allen distinguished two possible models for such a mechanism ("capacitor discharge" vs "neural counter") and discussed evidence for and possible consequences of each. For example, although Creelman (1962) writes that his data are incompatible with any periodic clock for temporal discrimination, thus arguing against a cyclically activated neural generator, both Michon (1967) and Kristofferson (1976) present data with distinctly periodic components. No direct comparison of the various models suggested so far for controlling speech timing has been performed, however, and the issue remains open.

This brief sampling of models of speech timing may be summarized as follows. (1) Most studies modeling timing in speech production either have described the temporal properties of known production units, such as segments, or have sought evidence of unknown units or the mechanisms whereby they are produced. (2) Although there have been some methodological differences among studies, their results have been in substantial agreement, at least within major classes of models. (3) Many important issues raised by these studies are apparently testable, but great care will be required to avoid methodological pitfalls.

What is the shape of things to come in this area of study? Will tomorrow's models be refined variants of today's, or will new concepts force a radical restructuring of our thought? The answer, I believe, is "both". For some purposes, such as practical speech synthesis, refinements and straightforward extensions of present descriptive models will be adequate for some time. Here the output must be acceptable as fluent speech, but the process by which it is generated need not model human (neuro-) physiology.

There is already under way, however, at least one radical restructuring, which will affect profoundly the form of models of speech production and perception. Turvey (1975) and several of his colleagues have argued persuasively for what they call an "action theory" of speech production, in which the motor system's normal reflexes are organized into ever higher levels of coordination, the highest level of all being sensibly describable only in terms of the overall goal, or plan, of the action. Such mainstays of traditional speech production research as "segment", "coarticulation", and "motor unit" become, in this view, projections of the plan onto subspaces of greatly reduced dimensionality, so simplified in most cases as to obscure the "true" process of production. Fowler (1977) has examined the implications of this kind of theory for models of speech timing, giving us a good opportunity to glimpse at least the immediate future.

At a rather deep level of conceptualization, we may see more explicit appeal to the goals of the speech timing model; that is, it will be not only acceptable but even necessary to consider the function of temporal structure in order to understand adequately what we observe. For example, such a statement as "Speech is made to be spoken" (Allen, 1975) will become literal rather than

figurative truth.

Models of "intrinsic timing", as Fowler (1977) terms them, may impose far more explicit constraints on the domain of control than do many present-day models. Since in that view the temporal figure is as much a part of the speech act as, say, its neuromuscular features, intrinsic timing is an inherent property of the act, coterminous with it, not something that is imposed on it by an external timing generator that exists before and after as well as during. Hence it would be improper to speak, for example, of "the effect of speaking rate on segment duration", since the effect is really on the whole structured act within which the segment is embedded.

Some models already refer explicitly to domains of temporal constraint. Lindblom and Rapp (1973), for example, use one parameter to describe the effect of the number of syllables following within the same word and a second for the number of words following within the same phrase. Allen (1973) restricts his model of timing control to effects within the breath group. Other local constraints, such as neighboring phonemic context, are commonly imposed. Even so we may soon find the focus changing in our consideration of domains of temporal constraint; since the timing is intrinsic to the act, we would seek either to isolate acts as delimiters of temporal domains or to identify differences in timing control as evidence of action boundaries. We have often done this before, but usually intuitively or even unconsciously, and with segmental phonology and orthography as our guides. Following "action theory" into hierarchical systems of coordinated reflexes may bring us some interesting surprises.

Finally, we should still find as much need in models of intrinsic timing as in our present models for the notions of "temporal compensation" and "timing control mechanism" ("clock"). The assumption that motor action plans are organized hierarchically implies that temporal compensation will appear at all levels below the very highest; otherwise the temporal figure could not be intrinsic to the plan. Moreover, as long as neuromuscular events within the plan do not follow rapidly one upon the other, as fast as the associated lowest level reflexes allow, a controlling mechanism must be assumed to decide when to move on to the next. It could be proposed that the neural structures and pathways responsible

for coordinating the action of muscles in space are simultaneously responsible for their temporal patterning as well, i.e., the plan is its own clock. The dissociation of temporal from spatial control in such dysrhythmic conditions as cerebellar ataxia, however, suggests strongly that a separate mechanism will continue to be needed in adequate models of timing in motor action plans.

In conclusion, we may expect that descriptive models of speech timing will continue to be elaborated, with fairly clear lines of historical development from the very earliest descriptions of segment durations. Theoretical models, on the other hand, may be about to undergo substantial modification, as we revise our conceptualization of the speech production process and of the relationship of timing to that process.¹

References

- Allen, G.D. (1973): "Segmental timing control in speech production", JPh 1, 219-237.
- Allen, G.D. (1974): "Measurement error in speech timing studies", JASA 55, Suppl. 1, S42 (abstract).
- Allen, G.D. (1975): "Speech rhythm: its relation to performance universals and articulatory timing", JPh 3, 75-86.
- Allen, G.D. (1978): "Vowel duration measurement: A reliability study", JASA 63, 1176-1185.
- Bernstein, N.A. (1967): The Coordination and Regulation of Movements, Oxford: Pergamon Press.
- Cooper, M.H. and G.D. Allen (1977): "Speech timing control in normal speakers and stutterers", JSHR 20, 55-71.
- Creelman, C.D. (1962): "Human discrimination of auditory duration", JASA 34, 582-593.
- Fowler, C.A. (1977): Timing Control in Speech Production, Indiana: University Linguistics Club.
- House, A.S. and G. Fairbanks (1953): "Influence of consonant environment upon the secondary acoustic characteristics of vowels", JASA 25, 105-113.
- Huggins, A.W.F. (1972): "Just noticeable differences in segment duration in speech", JASA 51, 1270-1278.
- Keating, P. and C. Kubaska (1978): "Variation in the duration of words", JASA 63, Suppl. 1, S56 (abstract).
- Kim, C.-W. (1966): "The linguistic specification of speech", UCLA: Working Papers in Phonetics, 5.
- Klatt, D.H. (1975): "Vowel lengthening is syntactically determined in a connected discourse", JPh 3, 129-140.

(1) The support of NSF grant number BNS 7614345 and NIH grant number RR 05333-16 is gratefully acknowledged.

- Kozhevnikov, V.A. and L.A. Chistovich (1965): Speech: Articulation and perception, Joint Publications Research Service, Washington, D.C., 30,543.
- Kristofferson, A.B. (1976): "Low-variance stimulus-response latencies: Deterministic internal delays?", Perc.Psych. 20, 89-100.
- Lehiste, I. (1972): "Timing of utterances and linguistic boundaries", JASA 51, 2018-2024.
- Lindblom, B. and K. Rapp (1973): "Some temporal regularities of spoken Swedish", Papers from the Institute of Linguistics, University of Stockholm.
- Michon, J.A. (1967): Timing in Temporal Tracking, Soesterberg, the Netherlands: Institute for Perception.
- Nooteboom, S.G. (1972): Production and Perception of Vowel Duration, Doctoral dissertation, University of Utrecht.
- Ohala, J.J. (1970): "Aspects of the control and production of speech", UCLA Working Papers in Phonetics 15.
- Ohala, J.J. (1975): "The temporal regulation of speech", in Auditory Analysis and Perception of Speech, G. Fant and M.A.A. Tatham (eds.), New York: Academic Press, 431-452.
- Peterson, G.E. and I. Lehiste (1960): "Duration of syllable nuclei in English", JASA 32, 693-703.
- Pike, K.L. (1945): The Intonation of American English, Ann Arbor: The University of Michigan Press.
- Tingley, B.M. and G.D. Allen (1975): "Development of speech timing control in children", Child Devel. 46, 186-194.
- Turvey, M.T. (1975): "Preliminaries to a theory of action with reference to vision", in Perceiving, Acting and Knowing: Toward an Ecological Psychology, R. Shaw and J. Bransford (eds.), Hillside, N.J.: Lawrence Erlbaum Associates.
- Umeda, N. and C.H. Coker (1975): "Subphonemic details in American English", in Auditory Analysis and Perception of Speech, G. Fant and M.A.A. Tatham (eds.), 539-564, New York: Academic Press.
- Wright, T.W. (1974): "Temporal interactions within a phrase and sentence context", JASA 56, 1258-1265.

THE EFFECT OF SENTENCE ACCENT ON QUANTITY

Robert Bannert, Phonetics Laboratory, Lund University, Sweden

This paper will focus on the phonological aspects of the temporal structure of quantity in Central Swedish. Its domain is the vowel and consonant sequence resulting in the temporal pattern of complementary length, namely vowel + short consonant (V:C) or short vowel + long consonant (VC:). Quantity will be studied from the sentence perspective by investigating the prosodic effect of sentence accent (SA) on the VC-sequences.

Investigation

The tonal manifestation of SA in Stockholm Swedish is treated exhaustively by Bruce (1977). The present investigation builds on his tonal findings. Two speakers from Stockholm, EH, female, the main informant in Bruce (1977) and TB, male, representing the same dialectal variety, produced the test material. The test words were stöka (V:C) and stöcka (VC:). The qualitative difference between the long and short vowel is rather small. They were placed alternatively in one of the three positions (1,2,3) in the base sentence "Man kan lämna långa nunnor efter åtta." (One can leave tall nuns after eight). The test words, like the basic words (underlined) in the three sentence positions, are stressed with word accent 2 on the first syllable.

The test material consisted of 12 sentences. Six of them contained the test words in one of the three accented positions without SA, SA falling on the adverbial. The other six sentences contained the test words with SA.

Results and discussion

The VC-sequences without sentence accent are regarded as a reference for studying the effect of SA on quantity.

1. The temporal structure of quantity. The overall means of the durations of the vowel and the following consonant are plotted in figure 1. The diagram represents the temporal space for the manifestation of the VC-sequences (lower limitations in terms of incompressibility, cf Lindblom et al 1976, left aside). The range is illustrated by ellipses, constructed according to the standard deviations of vowel and consonant. The data points of both speakers and both types of sequences fall in such a way that they may easily be accounted for by straight lines. The temporal structure

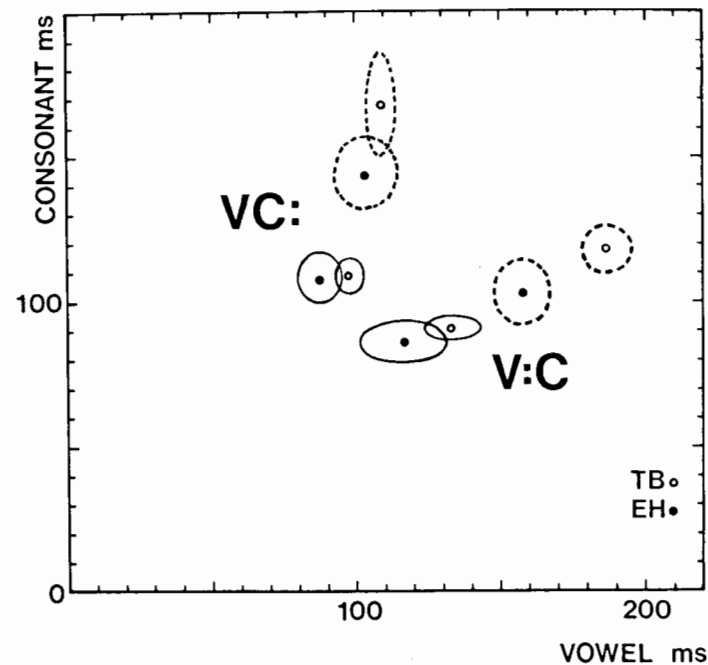


Figure 1. The effect of sentence accent on the temporal structure of the /V:C/- and /VC:/-sequences, consonant durations plotted against vowel durations. Sentence accent with broken-line circles. Pooled means and standard deviations for two speakers from Stockholm.

of the two types of sequences with and without SA are very similar for both speakers. Figure 1 shows also that the sequences with SA not only lie further apart from those without SA: Each type of sequence is also more separated from its counterpart.

The temporal effect of SA on quantity in Stockholm Swedish is a considerable increase of the segment durations which makes the temporal structure of the two contrasting VC-sequences more dissimilar. Thus the temporal contrast becomes clearer in focus position, given more prominence by the SA.

2. The increase of segment durations. The durations of the segments do not increase in a uniform way. It is evident from figure 1 that the increase of segment duration is largest for the long segment of each type of sequence, i.e. the long vowel in (V:C) and the long consonant in (VC:).

The V/C relations. The proportion of the increase of segment duration in each position and in all positions together is given in table 1. The ratio of the durational increase is defined as the relationship between the consonant and the vowel, $k = \Delta C / \Delta V$. It expresses the degree of lengthening of the consonant compared to that of the preceding vowel.

Table 1. Increase in segment durations. The factor k gives the proportion of the lengthening of the consonant compared to that of the preceding vowel.

sequence	speaker	position				pooled
		1	2	3	1-3	
V:C	TB	0.48	0.43	0.72	0.53	0.48
	EH	0.46	0.48	0.39	0.43	
VC:	TB	4.70	5.17	6.18	5.36	3.50
	EH	3.00	1.76	2.06	2.23	

Although there is some variation across the three sentence positions for both speakers, the short consonant is prolonged approximately by a factor of 0.5 of the preceding long vowel, while the duration of the long consonant increases much more in comparison with its preceding short vowel.

A non-uniform change of segment durations in VC-sequences with complementary length is also found in other prosodic contexts in Central Swedish and also in Central Bavarian, namely with differing speaking rates and stresses (contrastive vs neutral) and for words pronounced in isolation vs embedded in a sentence (cf Gårding et al 1975, Bannert 1976). This would suggest that the increase in segment duration due to different conditions is both linear and non-uniform.

The degree of lengthening. When the relationship of the non-uniform increase in segment duration between the vowel and the following consonant is established, it is sufficient to know the degree of lengthening for only one segment, e.g. the vowel. The relative degree of lengthening of the long and short vowel is given in table 2.

For this parameter, too, the change is not invariant. Both speakers behave differently for the three sentence positions and for the two categories of vowels. These differences may be accounted for with reference to two individual differences:

1) There are different durations between the speakers on the three

Table 2. Percentage of the increase in vowel duration with SA.

sequence speaker		position				
		1	2	3	1-3	pooled
V:C	TB	42.7	48.4	29.9	39.9	37.2
	EH	28.4	38.9	36.1	34.3	
V C:	TB	10.2	12.5	10.9	11.2	14.5
	EH	14.9	20.5	19.8	18.3	

sentence positions in the reference cases without SA (figure 1).

- 2) The speakers have two different tonal behaviours. Speaker EH has a tonal rise in the postconsonantal vowel, the pitch level of which is lower than that in the accented vowel. Speaker TB, however, reaches the high F_0 -level for the SA, which is higher than that in the accented vowel, in the consonant itself. Therefore, TB seems to need more time for producing the consonant.

But these minor differences in temporal behaviour remain well within the typical pattern of complementary length.

3. Preserving the temporal identity of the VC-sequences

It can be hypothesized that the change in the temporal patterns of the VC-sequences is governed by a dominant phonological principle, which is due to perceptual mechanisms: Preserve or sharpen the temporal contrast.

Two ways of increasing segment durations. There are two possibilities for the means by which segment durations in focus position may be increased:

- (1) equally to both segments either in absolute terms (ms) or as a percentage of the segment duration without SA,
- (2) differently to both segments. Then the question arises: How different and why?

In figure 2, two kinds of equal increase of segment duration and their effect on the temporal pattern of the sequences are illustrated. The reference points are the pooled means for the durations of the /V:C/- and the /VC:/-sequences without SA for both speakers taken together.

An absolute increase in the duration in ms, equal for both the vowel and the consonant, will result in shifting the VC-points along a straight line, corresponding to the slope $k=1$. As a consequence, however, the temporal structures of the two contrasting sequences will become more similar to each other. It is known

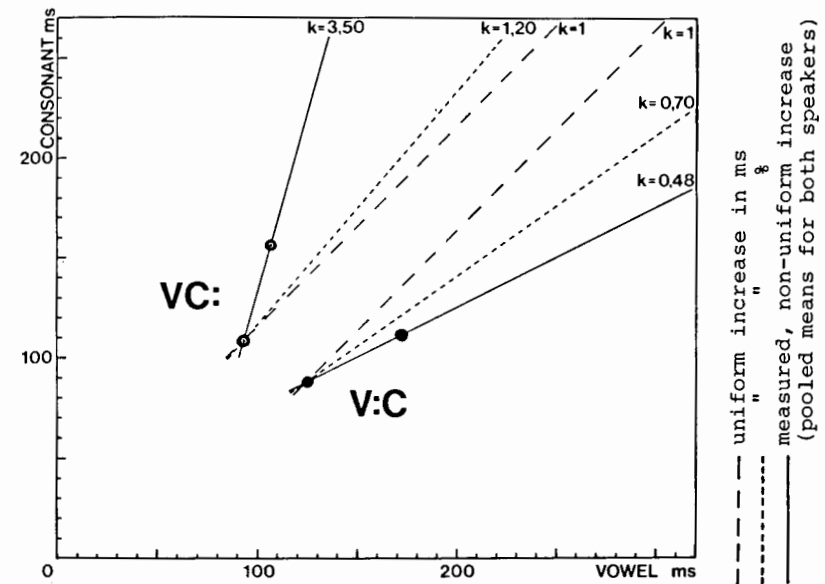


Figure 2. Different ways of increasing segment durations in the VC-sequences. The factor k gives the degree of lengthening for the consonant compared to the preceding vowel.

that durational differences must become greater with increasing segment duration in order to be perceived (cf the discussion in Lehiste 1970).

A relative increase of the duration, equal for both segments, is shown by the dotted lines. The slope (factor k) is now smaller for the /V:C/-sequence and larger for the /VC:/-sequence due to the asymmetric temporal structure within the sequences. This relative increase will result in enlarging the temporal difference between the sequences. But it seems obvious that this equal relative change of duration does not lead to a sufficient dissimilarity between the sequences with SA, either. The measured pooled means clearly lie further apart from each other than either of the possibilities would predict. Whereas the duration of the long vowel increases twice as much as that of the following short consonant ($k \approx 0.5$), the duration of the long consonant increases by far faster than that of the short vowel ($k \approx 3$).

The segment-to-sequence ratio. Due to the temporal patterns of the VC-sequences, they can be viewed as a unit of production and per-

ception. Then the relationship between the unit and its two parts will represent a measure of the temporal structure of quantity of complementary length. The vowel-to-sequence ratio $V/(V+C)$, expressing this relationship, was introduced by Bannert (1976) and applied to three languages with complementary length, Central Bavarian, Northern Icelandic, and Central Swedish.

Because of the phonological dependencies between the vowel and the following consonant, it seems inadequate to state the temporal structure of VC-sequences with complementary length by calculating the segment ratios $V/V:$ and $C/C:$ (cf the criticism in e.g. Lindblom et al 1976). Neither of these ratios can account for the temporal changes of the two speakers.

The segment-to-sequence relations, given as the $V/(V+C)$ ratios, are plotted in figure 3. It is clear that, when represented in this way, both speakers change the temporal structure of the sequences in exactly the same way.

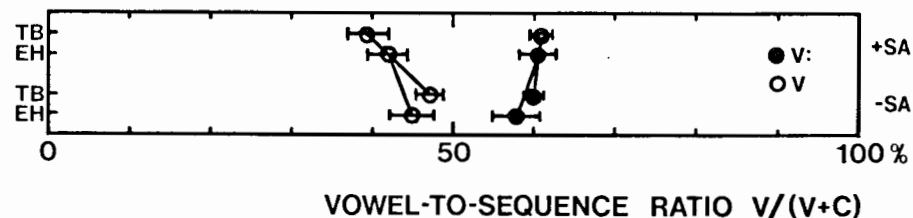


Figure 3. Variation of internal temporal structure of the VC-sequences in sentence accent position for each speaker. The relationship between the segments within the sequences is expressed by the vowel-to-sequences ratio.

The addition of SA to quantity in Stockholm Swedish increases the temporal distance, e.g. in terms of the segment-to-sequence ratio, between the two types of VC-sequences. Thus the temporal contrast for perception is well maintained or even enlarged.

References

- Bannert, R. (1976): Mittelbairische Phonologie auf akustischer und perzeptorischer Grundlage. Travaux de l'Institut de Linguistique de Lund X, B. Malmberg and K.Hadding (eds), Lund: Gleerup.
- Bruce, G. (1977): Swedish word accents in sentence perspective. Travaux de l'Institut de Linguistique de Lund XII, B.Malmberg and K.Hadding (eds), Lund: Gleerup.
- Gårding, E., O. Fujimura, H. Hirose, and Z. Simada (1975): Laryngeal control of Swedish word accent. Working Papers 10, 53-82. Phonetics Laboratory and Department of General Linguistics, Lund University.
- Lehiste, I. (1970): Suprasegmentals, Cambridge, Massachusetts: MIT Press.
- Lindblom, B., B. Lyberg, and K. Holmgren (1976): Durational patterns of Swedish phonology: Do they reflect short-term memory processes? Mimeographed. Department of Phonetics, Stockholm University.

SOME NOTES ON THE PERCEPTION OF TEMPORAL PATTERNS IN SPEECH

Rolf Carlson*, Björn Granström*, and Dennis H. Klatt, Mass. Inst. of Tech., Cambridge, MA 02139 USA. [*Also Dept. of Speech Communication, KTH, S-10044 Stockholm, Sweden.]

Introduction. Prosodic factors in speech have recently attracted a remarkable amount of linguistic and phonetic research. A prevalent point of view is that prosody is of paramount importance, both for naturalness and intelligibility of speech. As a result of this belief, a change can now be seen in the methods adopted in speech training for hard-of-hearing and foreign language students. The increased focus on suprasegmental compared to segmental articulation is possibly advantageous. From a scientific point of view, however, very little evidence is yet available on the quantitative importance of prosody. This is especially true of the relative importance of different aspects of the prosodic pattern.

From a study employing synthetic speech (Huggins, 1976), we know that really deviant durations and fundamental frequency contour decreases intelligibility. Prosodic parameters have also been shown to be effective in disambiguating sentences (Lehiste *et al.*, 1976). Our concern, however, has more to do with what information an explicit description of prosody has to supply and the precision with which it is supplied.

Descriptive models for segmental duration and fundamental frequency have been designed for a number of languages. Typically these models are based on material read repeatedly by a single speaker in a neutral, non-emphatic way. Subjects can perform remarkably consistently within such a recording session, but an examination of spontaneous speech reveals great variability in the prosodic realizations of a given sentence.

Thus it is not clear how precise the specification of duration is in the speech code common to speaker and listener. We also know that perception imposes certain restrictions on how prosodic effects could be appreciated (Klatt and Cooper, 1975). From previous studies (Carlson and Granström, 1975; Fujisaki, 1975), we know that the sensitivity to durational changes is greater in vowels than in consonants. The durational balance

between syllable nuclei, as well as the interval between onsets of stressed vowels (a measure related to the foot concept) have been shown to be perceptually important (Carlson and Granström, 1975; Huggins, 1972; Lehiste, 1977).

This leads to the questions that we wish to address: given a primary interest in the functional properties of a model of prosody, what demands should we put on it? What aspects of the description are most important? Will different models be ranked in the same order if different criteria such as naturalness and intelligibility are used?

In our present study we have evaluated both the naturalness and intelligibility of sentences with several different durational structures. As a starting point we have used a version of Klatt's durational rules for American English (Klatt, 1979) that we use in the MIT text-to-speech project (Allen, 1976).

Test Material. An algorithmically complete rule system is meant to generate a first order approximation to the durational structure of any spoken English sentence. In order to evaluate such a system of rules, a variety of syntactic and phonological structures ought to be tested. The test material in our experiments, presented below, could include only a small sample of such structures. These include the active, passive, question, simple, compound, and complex embedded sentence types. Both short and long noun phrases are represented. In Sentence 8, the ")n" after "seafood" is specially used to indicate that the following prepositional phrase is a sentential modifier, rather than modifying the "icy seafood" noun phrase.

Test Sentence	measured dur. (msec)	synth.dur. (msec)
1. Someone at the table)n ordered hot and sour soup.	2365	2615
2. Going to school)n was an adventure.	1625	1860
3. He who eats too much)c will become fat.	2105	2295
4. If Kate)n goes, Bill)n will eat her orange.	2430	2385
5. Old eggs)n often spoil french bread.	2200	2330
6. The fat brown turkey)n was chased by everyone.	2495	2415
7. Do you think that it will rain?	1370	1450
8. Pete)n ate icy seafood)n on the veranda.	2195	2155
9. Frank)n saw pretty streetcars)n in San Francisco.	2755	2845
where: Noun Phrase Boundary =)n, Clause Boundary =)c		

These sentences were recorded several times, the most natural sounding recording was selected, and the duration of each segment was measured. Since the rules are intended to match the speech of a particular speaker (DHK), the same subject was employed in the recording session. Nine different versions of each sentence were synthesized and put on language master cards. The synthesis algorithm is discussed in Klatt (1979). The versions listed below include three (3-5) that might be expected to be preferred over version Rule (since the Rule durations are adjusted in part toward Ref durations) and four versions (6-9) expected to be worse than Rule (since various rules contained in Rule have been deleted).

- 1 Ref Synthesis by rule using the measured durations from natural speech, but normalized linearly over the whole sentence to get the same total duration as Rule. This adjustment of the speech rate was rather small, averaging 6 percent.
- 2 Rule Synthesis using the rule system.
- 3 Vowel Synthesis using the vowel durations from Ref and the consonant durations from Rule.
- 4 Cons Synthesis using the consonant durations from Ref and the vowel durations from Rule.
- 5 StressVO Synthesis using the durations from Rule but linearly normalized between stressed vowel onsets to get the same durations between onsets of stressed vowels as in Ref.
- 6 NoParse Synthesis using the rule system, but disregarding the syntactic boundaries marked by ")n" and ")c".
- 7 SimpleFL Same as Simple (below) but with clause-final lengthening at punctuation marks. Each segment after and including the last stressed vowel is assigned increased duration by a factor of 1.65.
- 8 Simple Synthesis using a very simple rule system:
 Stressed vowel : Dur= .80 * inherent duration
 Unstressed vowel : Dur= .60 * inherent duration
 Stressed consonant : Dur= .90 * inherent duration
 Unstressed consonant : Dur= .65 * inherent duration
- 9 Random Synthesis using the reference duration, but randomly multiplied or divided by a factor determined by the deviation between Rule and Ref. This condition was included as a clear example of a bad system.

Experiment I: Naturalness. Nine phonetically trained subjects (native speakers of American English, working at RLE, MIT) were asked to sort the nine versions of each sentence according to naturalness of the durational structure. The subjects used a language master and headphones. After the order for a particular

sentence type was settled, the subject assigned a number corresponding to subjective naturalness (from 0 - 100) to each version. Most of the subjects finished the task within two hours. In some cases, the task required several sessions. Since the subjects used different scales in the rating task, the data from each subject were normalized to produce a mean of 0 and a standard deviation of 100. Mean ratings across sentences (Table 1, Column labeled "mean") indicate that Ref, Rule, Vowel, and StressVO are judged to be significantly more natural than the others.

The reproducibility of the naturalness rating for a subject was estimated from sentence seven, which had no syntactic markers in the input representation, and was thus identical in versions Rule and NoParse. The mean normalized distance across subjects for this pair was 26, which compares favorably to a typical standard deviation of 60 for the observations underlying an element in the matrix (Table 1). This suggests that subjects were quite consistent in their ratings compared to the intersubject variability. The estimated standard deviation of each element in the matrix is about 20 (60 divided by the square root of 9). The standard deviation of the mean across sentences is given in the table for each version.

Table 1. Naturalness ratings from Experiment I, as averaged across nine subjects. Column A indicates the number of errors for 6 versions used in an intelligibility test described in Experiment II. Versions 3, 4 and 9 were not included in Experiment II.

ver- sion	sentence									mean	st.d.	A
	1	2	3	4	5	6	7	8	9			
1	78	58	17	33	21	68	-18	51	78	43	8	2
2	18	33	47	32	70	70	86	17	23	44	7	12
3	86	60	28	0	52	104	44	81	40	55	7	-
4	0	-2	6	73	38	71	-25	-23	61	22	8	-
5	99	28	-2	81	35	108	-25	50	92	52	8	9
6	14	66	3	30	-101	33	70	11	54	20	8	12
7	-36	4	22	21	-114	4	43	6	-80	-14	9	15
8	-63	-66	3	-169	-146	-30	-125	13	-127	-79	11	24
9	-116	-168	-169	-132	-163	-148	-150	-190	-80	-146	9	-

Experiment II: Intelligibility. Some of the versions used in Experiment I were included in an intelligibility test that was presented individually to 18 MIT students. These subjects were phonetically naive, native speakers of American English, and unfamiliar with synthetic speech. Before the test was run, the subject listened to a short passage of synthetic speech (75 sec)

to get acquainted to the speech quality. This familiarization process has been shown to be very rapid (Carlson et al., 1976). The number of word errors out of 122 possible words (excluding articles) is shown in Column A of Table 1, and is plotted against the naturalness rating data in Fig. 1.

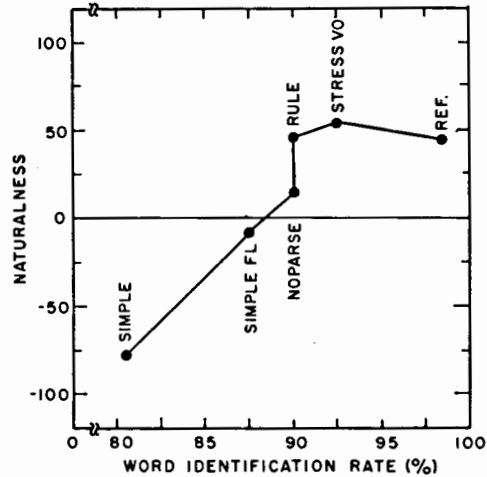


Figure 1. Mean naturalness ratings for six versions are plotted against the word identification rate from Table 1. The two measures are positively correlated, but the improvement in naturalness from NoParse to Rule is not accompanied by an improvement in intelligibility. It should be emphasized that the intelligibility figures are based on a small amount of data and should be interpreted with some caution.

Discussion. It is clear from ratings and comments given on the answer sheets, that subjects have different preferences. For example, Ref is not considered best by all subjects for all sentences. This might be a question of dialectal preference or idiosyncratic differences. Another possibility is that durations from natural speech, imposed on synthetic speech with a somewhat different realization of F0 and segmental content could constitute an incompatible combination. There is no way of controlling for this in the present study. A parallel study using LPC-coded natural speech might shed some light on this issue.

Ref and Rule have about the same mean naturalness score in Table 1, indicating that the durational rules produce as natural a durational structure as our reference speaker <POINT 1>. However, it should be noted that the test material consist of rather short sentences without e.g. the semantic relations between sentences that exist in paragraph-length material and that the intelligibility of Rule was somewhat lower than that of Ref.

We wanted to examine how the intermediate versions between Ref and Rule (Vowel, Cons, and StressVO) are ordered. This could not be done if we are not sure that the relation between Ref and Rule is the same for all subjects. Therefore, in Table 2, the results are presented after discarding the data on a sentence for each subject who rated Rule higher than Ref. (This will, of course, reduce the naturalness score for Rule relative to all other versions.)

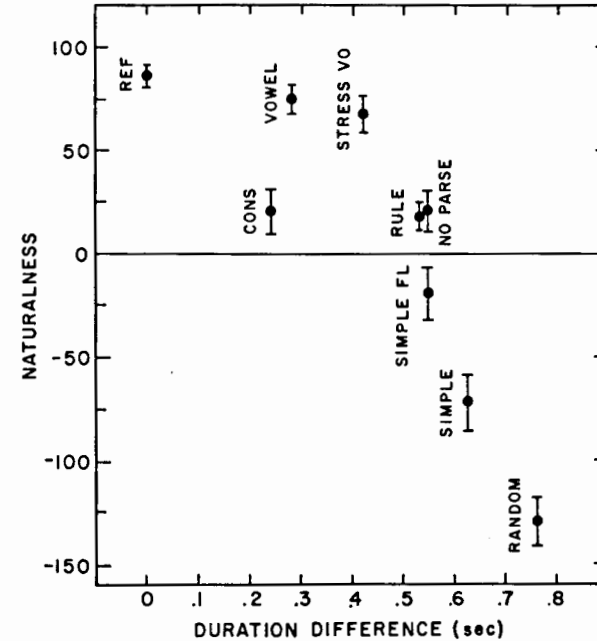


Figure 2. Mean naturalness ratings from Table 2 are plotted against one measure of the physical durational distance to Ref (city block), i.e., the sum, over all segments, of the absolute difference in duration between the version and Ref (average per sentence). There is a general correlation between naturalness ratings and physical difference in duration, but Rule, StressVO and Vowel are rated more natural than one might expect given the durational differences involved.

Table 2. Naturalness ratings after excluding Ss who rated Rule better than Ref. The number of subjects for each sentence is marked in the last line.

ver- sion	sentence									mean	st.d.
	1	2	3	4	5	6	7	8	9		
1	102	85	66	55	73	102	93	96	83	85	6
2	13	12	10	4	59	42	44	14	12	18	8
3	110	85	78	6	73	116	72	82	40	74	8
4	2	-20	-42	92	90	91	25	-43	52	20	11
5	105	27	27	53	42	116	41	56	90	67	9
6	-4	43	-30	52	-125	37	44	19	56	20	10
7	-40	-10	36	26	-159	-11	17	13	-82	-20	13
8	-42	-68	-27	-176	-208	-30	-71	-4	-119	-73	14
9	-125	-133	-150	-121	-122	-150	-116	-180	-75	-130	12
	7	6	5	5	2	5	2	7	8		(# subjects)

The most striking result seen in Figure 2 is that both Vowel and StressVO are significantly more natural than Cons, despite their greater durational distance from Ref. This corroborates earlier observations that these two durational units i.e. vowel duration and interval between onset of stressed vowels are of great perceptual importance <POINT 2>. Cons, which has all consonant durations right but vowel durations done by rule, does not score significantly better than Rule, reinforcing this interpretation. Furthermore it is obvious that physical distance is clearly not a reliable predictor of perceptual distance.

Isochrony, i.e. the tendency toward equal durations between certain units, has been discussed in the literature. It might be suspected that the high scores for Ref and StressVO are because they preserve the isochrony of real speech. We compared Rule and Ref to see which has a greater tendency toward equal distances between stressed vowel onsets. If anything, Rule is more isochronous than Ref, suggesting that the amount of isochrony implemented in the rules via, e.g., cluster shortening and unstressed segment shortening is probably sufficient, and no "isochrony rule" per se need to be added <POINT 3>.

Versions Rule and NoParse have the same naturalness score in Figure 2. However, it must be remembered that an editing has taken place which selectively lowers the score of Rule. In Table 1, these two versions are significantly different. For some sentences, however, the score for NoParse is higher than that for Rule. Even if these differences are not highly significant, they indicate that in these instances, NoParse is regarded as close in quality to Rule. One possible reason could be that the rules dealing with phrase final lengthening overexaggerate the lengthening effect. An analysis yielded no support for this interpretation in our data. Another possibility which seems more reasonable is that the phrase-final lengthening rule is applied too frequently <POINT 4>. A simple-minded cure might be to ensure that short phrases containing only one content word are not affected by phrase final lengthening although recent work by Cooper et al (1978) indicates the likelihood of a more complex relation between surface structure and lengthening.

Comparing Simple and Rule, we can conclude that rules modifying the duration of a segment as a function of syntax and segmental context are of significant importance for both naturalness and intelligibility. Approximately half of the difference between the two versions seems to be explained by the extremely simple clause final lengthening rule used for SimpleFL <POINT 5>.

The intelligibility results shown in Figure 1 indicate a clear correlation between intelligibility and naturalness. Correct durations result in significantly better intelligibility and naturalness. This confirms in part the current belief in the importance of prosody to sentence perception. <POINT 6>.

References

- Allen, J. (1976), "Synthesis of Speech from Unrestricted Text", *Proc. IEEE* 64, 433-442.
- Carlson, R. and Granström, B. (1975), "Perception of Segmental Duration", in *Structure and Process in Speech Perception*, A. Cohen and S.G. Nooteboom (Eds.), Springer-Verlag, Berlin, 90-104.
- Carlson, R., Granström, B., and Larsson, K. (1976), "Evaluation of a Text-to-Speech System as a Reading Machine for the Blind", *STL QPSR* 2-3/1976, 9-13.
- Cooper, W.E., Paccia, J.M., and Lapointe, S.G. (1978), "Hierarchical Coding in Speech Timing", *Cognitive Psychology* 10, 154-177.
- Fujisaki, H., Nakamura, K., and Imoto, T. (1975), "Auditory Perception of Duration of Speech and Non-speech Stimuli" in *Auditory analysis and perception of speech*, G. Fant and M. Tatham (Eds.), Academic Press, London.
- Huggins, A.W.F. (1972), "On the Perception of Temporal Phenomena in Speech", *J. Acoust. Soc. Am.* 51, 1279-1290.
- Huggins, A.W.F. (1976), "Speech Timing and Intelligibility", *Proc. Attention and Performance* 7, J. Reguin (Ed.).
- Klatt, D.H. (1979), "Synthesis of Segmental Durations in English Sentences", *9th International Congress of Phonetic Sciences*, Copenhagen.
- Klatt, D.H. and Cooper, W.A. (1975), "Perception of Segment Duration in Sentence Contexts", in *Structure and Process in Speech Perception*, A. Cohen and S.G. Nooteboom (Eds.), Springer-Verlag: Heidelberg.
- Lehiste, I. (1977), "Isochrony Reconsidered", *J. Phonetics* 5, 253-263.
- Lehiste, I., Olive, J.P., Streeter, L.A. (1976), "The Role of Duration in Disambiguating Syntactically Ambiguous Sentences", *J. Acoust. Soc. Am.* 60, 1199-1202.

THE PERCEIVED RHYTHM OF SPEECH

Andrew Donovan and C.J. Darwin, Laboratory of Experimental Psychology, and Centre for Research in Perception and Cognition, University of Sussex, Brighton, England

Introduction

Attempts to model the duration or rhythm of the segments of connected speech fall into two broad groups (see Fowler, 1977, for a review). On the one hand are those which allow the syntactic structure of the utterance to perturb a segment's ideal duration (e.g. Lindblom & Rapp, 1973; Klatt, 1975) but which recognise no overall rhythmic patterning; on the other hand are those which allow an overall rhythmic pattern to be perturbed by limits on the segmental durations which must be compressed or expanded into that pattern (Abercrombie, 1964; Witten, 1977). This latter family of models has taken for its underlying rhythm a sequence of isochronous beats occurring on adjacent stressed syllables marking out rhythmic units called feet. The choice of an isochronous foot is based partly on linguistic intuition ("English utterances may be considered as being divided by the isochronous beat of the stress pulse into feet of (approximately) equal length". Abercrombie, 1964), and partly on the observable phonetic fact that syllables tend to be shorter, the more there are in a foot (Huggins, 1975; Fowler, 1977). In choosing between these two approaches a crucial question is the status of the isochronous beat. It is clearly not a phonetic fact since a foot with many syllables tends to be longer than one with fewer (Halliday, 1967; Allen, 1975). Is isochrony then a significant linguistic insight, or merely a poetic fiction? Lehiste (1973; 1977) has argued that the discrepancy between the linguistic intuition and the phonetic data may be due to a perceptual illusion. Perhaps we hear speech as more isochronous than it actually is. Indeed such an illusion is precisely what we would expect if perception undid those perturbations required by segmental constraints on an underlying regular rhythm, presenting to the listener the underlying rhythm of the speaker. Evidence for the perceptual reality of isochrony would thus argue for its inclusion in models of speech production.

Experiments

Our experiments extend the earlier observations by Lehiste (1973) and Coleman (1974) on listeners' inability to perceive the rhythm of speech veridically. We have used two tasks, a rhythm

matching task and a tapping task. The first two experiments used the rhythm matching task. Subjects adjusted the times between four noise bursts to match the overall rhythm of either a sentence or a control sequence of non-speech sounds. They could listen to the sound whose rhythm they were to match or the adjustable noise burst sequence by pressing one or other of two buttons; thus they could not hear the two stimuli simultaneously but were able to listen to each separately as many times as they liked while making the adjustments.

The sentence used in Experiment 1 was "A bird in the hand is worth two in the bush", synthesized on PAT from parameters derived from real speech. The noise burst sequence that subjects adjusted had four strong bursts corresponding to the four stressed syllables with appropriate intervening weaker bursts representing the unstressed syllables. By adjusting either of three knobs subjects could adjust the interval between adjacent stressed bursts, but the rhythm of the intervening weaker bursts was kept constant, scaled in tempo to the new inter-stress interval. Subjects matched the rhythm of two versions of the sentence; one had the natural pitch contour, the other a monotone. They also matched the rhythm of a control sequence of tones whose onsets were at the same time intervals as the stressed syllables of the sentence. Subjects performed seven matches (the first two of which were not analysed) to each of the three stimuli. The control was always done first followed by the two speech conditions in an order counter-balanced between subjects.

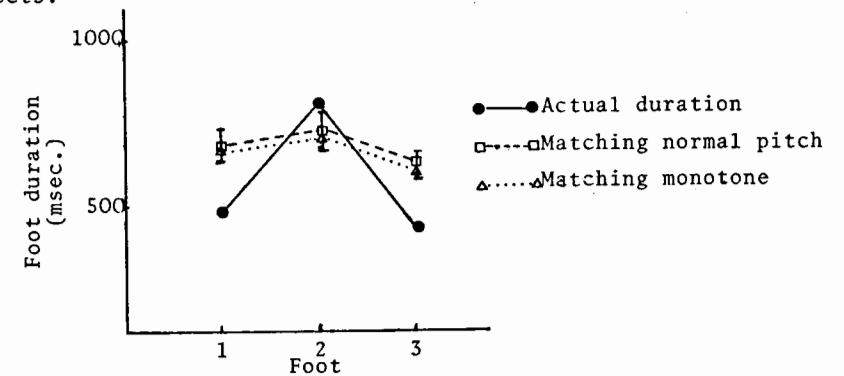


Fig.1. Actual and perceived foot durations for utterance in Experiment 1. (vertical bars represent \pm 1S.E. of the mean)

The actual and mean matched durations of the first three feet (that is the intervals between the four stressed syllables) are shown in Figure 1. To test any tendency for the matched durations to be more isochronous than the original utterance the quantity: $|(1-a_i/a_{i+1})| - |(1-p_i/p_{i+1})|$ where $a_i =$ actual duration of i^{th} foot
 $p_i =$ matched " " " " " " was calculated for $i=1,2$. A positive value for this quantity indicates that the perceived durations are more isochronous (over the two feet in question) than the actual durations. Such a tendency towards perceptual isochrony was reliable ($p<.001$) for both foot-ratios when subjects matched the two sentences, but was not found when they matched the control, non-speech rhythm. The natural and the monotone speech are both perceived as more isochronous than they really are, but the non-speech tonal pattern is not.

The second experiment differed from the first as follows:

- 1) Four sentences of natural (female) speech were used which contained different numbers of syllables in each foot.
- 2) The stressed syllables in each utterance all began with a stop consonant (/t/) and there were no other occurrences of this sound in the utterance. This made it easier to specify to the subjects where the major stresses fell as, instead of saying match the rhythm of the 'syllable beats' or the 'tapping points', they could be told to hit the /t/'s.
- 3) The noise-burst sequence was made up of five bursts only; an initial low amplitude burst corresponding to the first, unstressed syllable in each utterance, and four 'stressed' bursts corresponding to the four stressed syllables.
- 4) Subjects were explicitly encouraged to use a strategy that we had observed in the first experiment, namely repeating the sentence to oneself while listening to the noise bursts. In case subjects' own articulations were more isochronous than the original, recordings were made of each subject speaking each sentence. In fact we found they were not more isochronous and the following results still hold if matched durations are compared with subjects' own productions.

For the four sentences as a whole, four of the eight foot-ratios gave a significant ($p<.01$) tendency towards perceived isochrony, three gave no difference (partly because their foot ratios were actually quite close to unity already) and one gave a significant tendency away from perceptual isochrony. The results from

this deviant sentence and from one of the others are shown in Figure 2. Notice first in the right-hand panel that although subjects' judgements are quite reliable they are massively inaccurate at judging the duration of the middle foot. It is not clear though whether this huge overestimation of the middle foot should be attributed to perceptual isochrony. If it were, then we would not expect the similar, though more variable overestimation of the middle foot found in the left-hand panel for a sentence whose middle foot is already relatively long. Alternative explanations, which could account for the data from all four of the sentences, are that subjects overestimate the length of a foot containing (a) a major syntactic boundary or (b) a tone group boundary. Our third experiment looks at this question.

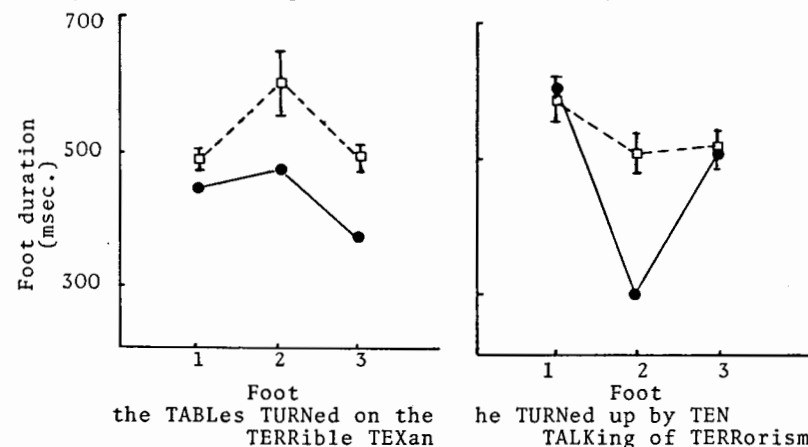


Fig.2. Actual and perceived foot durations for two of the utterances in Experiment 2.

To investigate the possible contribution of intonation to perceived rhythm a change of experimental technique was required. Subjects' imitations of sentences were, as we have seen, no more isochronous than the originals, but they did differ markedly in intonation. To ensure that subjects matched a sentence with the original intonation we asked them to tap in time to a sentence.

This new task differs from that used by Allen (1970) in that subjects tapped to every stressed syllable rather than just to a selected one on each trial. Here we are interested in subjects' perception of the entire rhythmic pattern. Subjects were not explicitly told to tap on the stressed syllables so the fact that they did provides an objective verification of the notion of the

rhythmic foot. The three utterances, which differed in number of tone groups and in syntactic structure, but which had identical foot durations were:

- 1) //1 Tim's in / Tuscany's / Training / Troops //
- 2) //1 Tim's in / Tuscany / Training / Troops //
- 3) //1 Tim's in / Tuscany //1 Training / Troops //

Here, following Halliday (1967), we bound tone groups by double slashes and indicated the type of tone group by a number. Both 2) and 3) contain a major syntactic boundary in the middle foot but in utterance 2) this was not marked by a tone group boundary. 1) and 2) were acoustically identical except that the /s/ of "Tuscany's" was spliced out for sentence 2) and replaced by four additional pitch periods of /v/ and an appropriate amount of silence to maintain the same foot length.

Fifteen subjects were divided into three groups, each group receiving a different order of presentation of the three utterances. Subjects heard each utterance 15 times and were told to start tapping after the third token. Each utterance was preceded by a warning tone 750 msec. from the onset of the utterance. Only the last 10 trials in each condition were analysed. Before each block of 15 trials, subjects heard the utterance three times and were given a context in which the utterances could occur. For example 3) might be the response to the question "Where's Tim and what's he doing?", while the same utterance with one tone group (sentence 2) might be the response to the question "What's Tim doing with the troops in Tuscany?" This was done to ensure that the subjects had a good idea of the syntax and tone group structure of the sentences they were listening to.

The results of this experiment (Figure 3) showed that while the number of tone groups has a distinct effect on perceived rhythm, the syntactic structure does not. In particular we found no tendency towards perceived isochrony in sentence 3, which contained two tone groups, but we did find a significant ($p < .01$) tendency towards perceived isochrony for both foot-ratios in sentence 1 and 2. Sentences 1 and 2 did not differ from each other significantly in this respect, but both differed significantly from sentence 3 ($p < .01$).

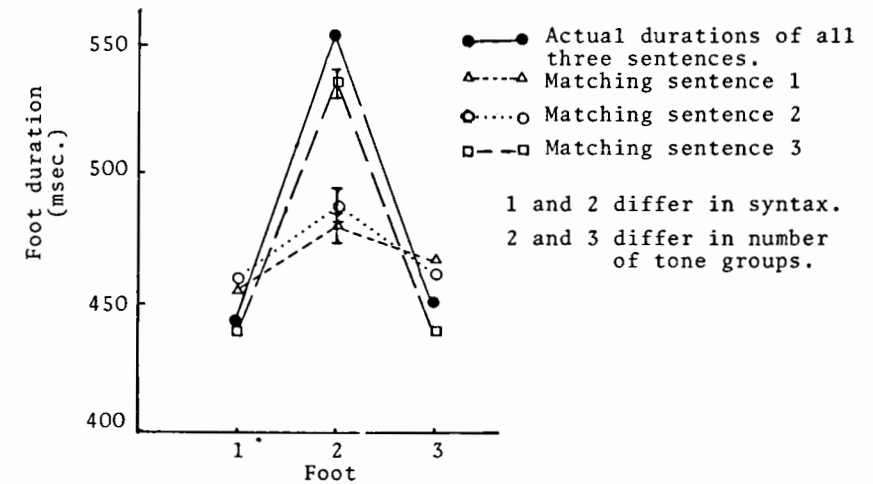


Fig. 3. Actual and perceived foot durations for the three utterances in Experiment 3 (see text for details).

It is apparent from these results that subjects are responding differently to the one and two tone-group utterances irrespective of the syntactic structure and despite the fact that the foot durations are the same in all three cases. Rees (1975), building on Halliday's (1967) work, has proposed that the tone group is a unit of rhythm as well as a unit of intonation so that isochrony need be maintained within but not between tone groups; it may be that this puts constraints on the limits of perceptual isochrony as well as on the tendency towards isochrony in production. It is clear from the experiments reported here that people are consistently inaccurate when judging speech rhythms and, furthermore, that they tend to hear these rhythms as more regular than they really are, at least when the utterance is bounded by a single tone group. Within the tone group, long feet tend to be underestimated, even when they contain a major syntactic boundary, while short feet tend to be overestimated.

Conclusions

Our results have broadly confirmed Lehiste's proposal that isochrony is partly a perceptual phenomenon. But we would make two points in addition. First, it is a perceptual phenomenon which is not independent of intonation. Second, we feel that it is a perceptual phenomenon, confined to language, reflecting underlying processes in speech production. Our results strengthen the case for models of the timing of English that incorporate an underlying

rhythmic organisation within tone groups. Conversely, they question the value of seeking direct links between syntax and segmental durations rather than indirect ones via an overall rhythmic structure which is also determined by the pragmatic and semantic context of a sentence (cf. Cutler & Isard, in press).

References

- Abercrombie, D. (1964): "Syllable quantity and enclitics in English", in In Honour of Daniel Jones, D. Abercrombie, D.B. Fry, P.A.D. MacCarthy, N.C. Scott, J.L.M. Trim (eds.) London: Longmans, 216-222.
- Allen, G. (1970): "The location of rhythmic stress-beats in English: An experimental study", UCLA Working Papers 14, 80-132.
- Allen, G. (1975): "Speech rhythm: its relation to performance universals and articulatory timing", JPh 3, 75-86.
- Coleman, C. (1974): A study of acoustical and perceptual attributes of isochrony, Ph.D. thesis, Univ. Washington.
- Cutler, A. and S.D. Isard (in press): "The production of prosody", in Language Production, B. Butterworth (ed.) New York: Academic Press.
- Fowler, C.A. (1977): Timing Control in Speech Production, Ph.D. thesis, Univ. Connecticut, Connecticut, Ind: Indiana Univ. Linguistics Club.
- Halliday, M.A.K. (1967): Intonation and Grammar in British English, The Hague: Mouton.
- Huggins, A.W.F. (1975): "On isochrony and syntax", in Auditory Analysis and Perception of Speech, G. Fant and M.A.A. Tatham (eds.), London: Academic Press, 455-464.
- Klatt, D. (1975): "Vowel lengthening is syntactically determined in a connected discourse", JPh 3, 129-140.
- Lehiste, I. (1973): "Rhythmic units and syntactic units in production and perception", JASA 54, 1228-1234.
- Lehiste, I. (1977): "Isochrony reconsidered", JPh 5, 253-263.
- Lindblom, B. and K. Rapp (1973): "Some temporal regularities of spoken Swedish", Papers from the Institute of Linguistics, University of Stockholm.
- Rees, M. (1975): "The Domain of Isochrony", Edinburgh Univ. Dept. of Linguistics, Work in Progress, 8, 14-28.
- Witten, I.H. (1977): "A flexible scheme for assigning timing and pitch to synthetic speech", L & S 20, 240-260.

TEMPORAL ORGANIZATION OF SEGMENTAL FEATURES IN JAPANESE DISYLLABLES

Hiroya Fujisaki and Norio Higuchi, Faculty of Engineering,
University of Tokyo, Tokyo, Japan

Introduction

While it is apparent that the realization of successive units in connected speech is based on the proper timing of articulatory and phonatory events, much remains to be investigated regarding the nature of the timing mechanism. It is not even agreed whether the regularity of timing (isochrony) resides in speech production or in speech perception, as pointed out by Lehiste (1977). The lack of our knowledge on this issue may primarily be due to the fact that the speech signal often fails to display marked segment boundaries, and that even the apparent boundaries do not directly reveal the timing of production nor the timing of perception. Elucidation of the mechanism underlying the isochrony thus requires experimental techniques for extracting, from the speech signal, the indices for the timing of production as well as the indices for the timing of perception of each of the successive units.

The present paper deals with both the productive and the perceptual aspects of the segmental timing in Japanese disyllabic words consisting only of vowels. Disyllabic words were selected since they display the characteristics of connected speech on the smallest scale. Vowel sequences were chosen since their acoustic characteristics can be most clearly defined in terms of formant frequencies, and the articulatory transition from the initial vowel to the second vowel can be traced in the trajectories of their formant frequencies.

The Speech Material

The speech material consisted of 20 disyllables, i. e., all the possible pairs of the five Japanese vowels (/i/, /e/, /a/, /o/, and /u/), pronounced with the "flat-type" word accent. Among these disyllables, nine were meaningful with the given accent type, four were meaningful when pronounced with a different accent type, and the rest were nonsense words. A randomized list of 100 words, containing five tokens each of the 20 disyllables, was read by a male speaker of the Tokyo dialect of Japanese. These disyllables were pronounced in isolation at an interval of three seconds. The speech signal was sampled at 10 kHz with an accuracy of 11 bits/sample and stored in the magnetic tape memory of a digital computer.

Analysis of Segmental Timing at the Level of Speech Production

An LPC analysis was made of all the utterances to extract the frequencies and bandwidths of 11 poles, from which the first three formant frequencies were selected on the basis of bandwidth and the continuity of the trajectories. These trajectories were then used to estimate the onset of the transition from the initial to the second vowel.

The estimation was based on the model of the coarticulation process in connected vowels previously proposed by Fujisaki et al. (1974, 1977). As shown in Fig. 1, the entire production process for connected vowels is represented by a hypothetical linear system which converts the stepwise target formant frequencies of each vowel into actual formant trajectories. An analysis of observed formant trajectories has indicated that a good approximation can be obtained by a critically-damped second-order linear system.

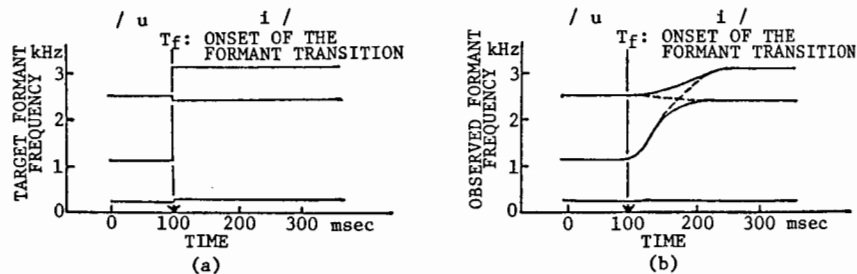


Fig. 1. Formulation of the process of coarticulation in the formant frequency domain: conversion of idealized formant target (a) into actual formant trajectories (b).

In the case of the disyllables under study here, we may assume a target frequency for the \$n\$th formant as

$$C_n(t) = F_{n1} + (F_{n2} - F_{n1}) u(t - T_f),$$

where \$F_{ni}\$ denotes the target frequency of the \$n\$th formant of the \$i\$th vowel, and \$T_f\$ denotes the onset of the transition measured from the voice onset of the initial vowel as the origin of the time axis. Then the actual formant frequency can be given by

$$F_n(t) = F_{n1} + (F_{n2} - F_{n1}) \{ 1 - (1 + \frac{t - T_f}{\tau_n}) \exp(-\frac{t - T_f}{\tau_n}) \} u(t - T_f),$$

where \$\tau_n\$ denotes the time constant for the transition of the \$n\$th formant. Further considerations regarding the continuity and cou-

pling of the resonance modes lead to good approximations of the formant trajectories for all of the vowel combinations. When a set of observed formant trajectories (\$F_1(t)\$, \$F_2(t)\$, and \$F_3(t)\$) is given, it is possible, by the method of Analysis-by-Synthesis, to determine the common onset of the formant transition and the time constants of individual formant trajectories. In the following analysis, a common time constant \$\tau_2\$ was assumed for the second and third formants. Examples of the observed formant trajectories and their best approximations by the above-mentioned model are shown in Fig. 2 for /ui/ and /iu/, where the estimated onset \$T_f\$ of the formant transition is also indicated.

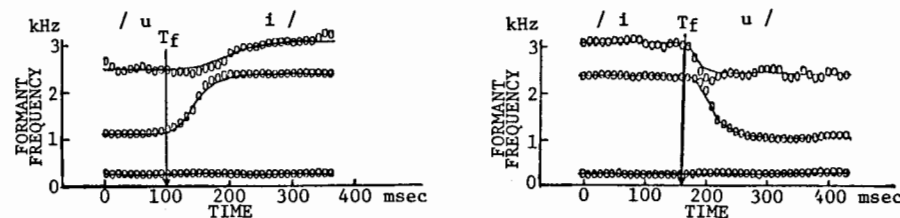


Fig. 2. Observed formant frequency trajectories (dots), their best approximations (—), and the estimated onset of the formant transition (\$T_f\$) for /ui/ (left) and /iu/ (right).

Table 1 summarizes the results for all the utterance samples and lists the mean values of \$T_f\$ and \$\tau_2\$ for five tokens of each disyllable. The following comments can be drawn from a comparison of the results for pairs of disyllables having the same vowel combination in a different order.

first vowel	second vowel				
	/i/	/e/	/a/	/o/	/u/
/i/	\$T_f\$	155	131	136	134
	\$\tau_2\$	21	25	22	27
/e/	\$T_f\$	90	132	134	149
	\$\tau_2\$	59	40	26	20
/a/	\$T_f\$	119	125	124	125
	\$\tau_2\$	48	39	41	37
/o/	\$T_f\$	101	94	92	113
	\$\tau_2\$	44	58	51	-
/u/	\$T_f\$	108	117	126	146
	\$\tau_2\$	39	44	28	-

Table 1. Mean values for the interval (\$T_f\$[msec]) from voice onset to the onset of formant transition and for the time constant (\$\tau_2\$[msec]) of the second formant trajectory for the 20 disyllabic words.

(1) In disyllables involving jaw movement without a change in lip articulation (i. e., /ie/ vs. /ei/, /ea/ vs. /ae/, /ia/ vs. /ai/, and /uo/ vs. /ou/), T_f is always larger for the disyllable produced by an opening movement of the jaw than for that produced by a closing movement. Analysis of variance indicates that the difference is highly significant (0.1% level) in /ie/ vs. /ei/, and is also significant (1% level) in /uo/ vs. /ou/, as well as in /ia/ vs. /ai/.

(2) In disyllables involving changes in lip articulation with or without minor jaw movement (i. e., /iu/ vs. /ui/, /eu/ vs. /ue/, /io/ vs. /oi/, /eo/ vs. /oe/, and /ao/ vs. /oa/), T_f is always larger for the disyllable produced by a rounding of the lips than for that produced by an unrounding of lips. The difference is significant in /eu/ vs. /ue/ (1% level); /ao/ vs. /oa/ (2% level); /eo/ vs. /oe/ (2% level); /io/ vs. /oi/ (5% level); and /iu/ vs. /ui/ (5% level).

(3) No significant difference in T_f was found for /au/ vs. /ua/, which involve both major jaw movement and changes in lip articulation in the transition from the initial to the second vowel. The effects of these two articulatory factors are considered to counteract and cancel each other.

These points may be easily observed in Fig. 3, which schematically shows the regions of the vowel target on the $F_1 - F_2$ plane. An arrow from one vowel target to another corresponds to a disyllable,

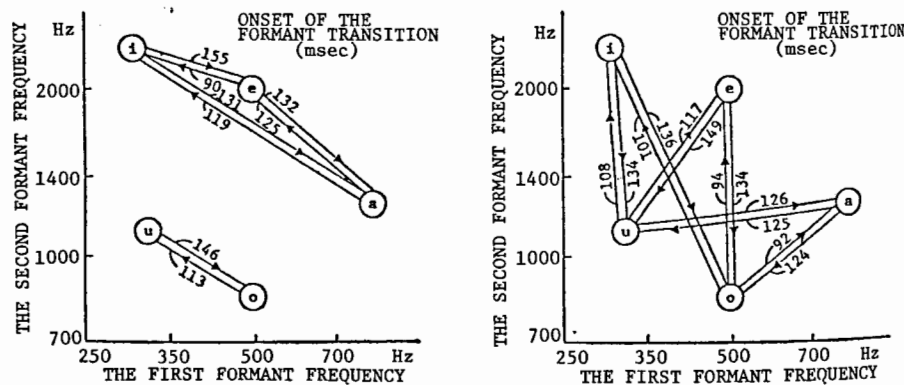


Fig. 3. Direction of the formant transition in the first and the second formant frequency plane and the onset of the formant transition (T_f).

and the number associated with the arrow indicates the mean value of T_f (in msec) for that disyllable.

Furthermore, there exists a very high negative correlation between T_f and τ_2 ($r = -0.91$) as shown in Fig. 4. Hence,

(4) Differences in the onset of transition (T_f) tend to compensate for the differences in the rate of transition; a slower transition is initiated earlier and vice versa.

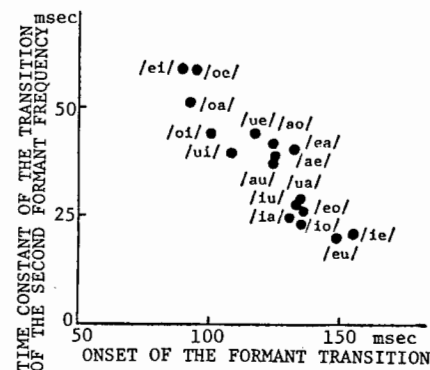


Fig. 4. Relationship between the onset of the formant transition (T_f) and the time constant (τ_2) of the second formant trajectory.

Analysis of Segmental Timing at the Level of Speech Perception

The last finding of the preceding section suggests the possibility that the apparent diversity in the onset of transition in various disyllables is introduced to maintain the uniformity of the perceived duration of segments. The following experiment was designed to investigate this possibility, using the same utterance samples as in the above analysis to find the instant of the perceptual onset of the second vowel within a disyllable.

A set of 20 points were selected at intervals of 5 msec to cover the range of the major formant transitions in the waveform of each disyllabic utterance. Twenty tokens of truncated disyllables were then prepared by curtailing the original speech waveform at these 20 points. These tokens were arranged in serial order at an interval of 3.5 sec as stimuli in an identification test using the method of limits. The subject was asked to answer whether he heard one vowel segment or two in a truncated disyllable. The test was repeated to obtain the response probability, and the perceptual onset of the second vowel was defined as the point corresponding to an equal probability for the two alternatives. An example of the stimuli and the subject's response probability is schematically il-

illustrated in Fig. 5. The test was conducted using one utterance of each of the twenty disyllables. The subjects were two male speakers of the Tokyo dialect.

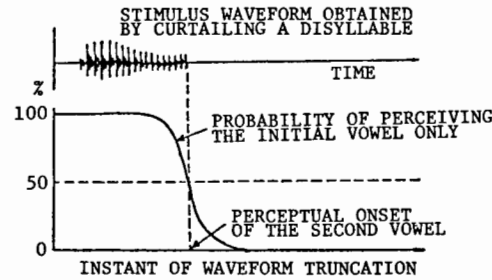


Fig. 5. Determination of the perceptual onset of the second vowel in a disyllable by waveform truncation.

Figure 6 shows the relationship between the perceptual onset (T_p) of the second vowel and the onset of formant transition (T_f) for each of the disyllables. Both T_p and T_f are expressed by their values relative to the total duration of an utterance. While the T_f 's for the various disyllables are distributed over a very wide range (22% - 42%), the T_p 's are found to be concentrated within a rather narrow range around the center of each utterance (48% - 53%). These findings suggest that the apparent diversity in the onset of the second vowel at the level of speech production may be the consequence of the speaker's effort to maintain the uniformity of perceived syllabic durations regardless of vowel combinations.

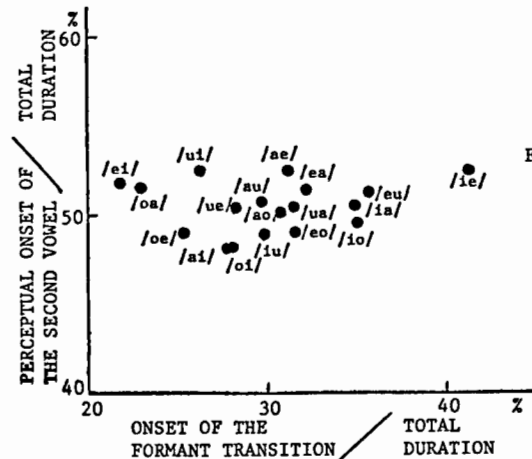


Fig. 6. Relationship between the perceptual onset (T_p) of the second vowel and the onset of the formant transition (T_f) for one sample of each of the disyllables.

Discussion

Two models of the possible mechanisms underlying the temporal organization of speech have been presented by Kozhevnikov and Chistovich (1965) and have since been widely discussed, e. g. by Ohala (1970), Leanderson and Lindblom (1972), and others. One is the so-called "chain model" based on the hypothesis of a closed-loop control of the speech production process. The other is the so-called "comb model" based on the hypothesis of an open-loop control. From our present knowledge concerning the motor organization of skilled behaviors, the chain model may be discarded, although it may certainly be true that various modes of feedback are necessary for the formation of the motor program. The findings of our present study suggest, however, that the comb model requires further elaboration. Our findings suggest that the formulation of the temporal relationship between the motor control and the articulatory/acoustic realizations of speech units is not complete without a consideration of their relationship to perceptual timing. From this point of view, two possible models can be distinguished under the open-loop (or "comb") hypothesis, as shown in Fig. 7.

In model (a), successive segments are produced with an isochronism at the level of the motor commands, so that their articulatory/acoustic realizations are not necessarily isochronous because of differences in the physiological and physical properties of the various articulators, as well as in the manner of articulation. In model (b), on the other hand, the motor commands and the articulatory/acoustic realizations of successive segments are programmed in such a way that the perceptual onsets of successive segments occur

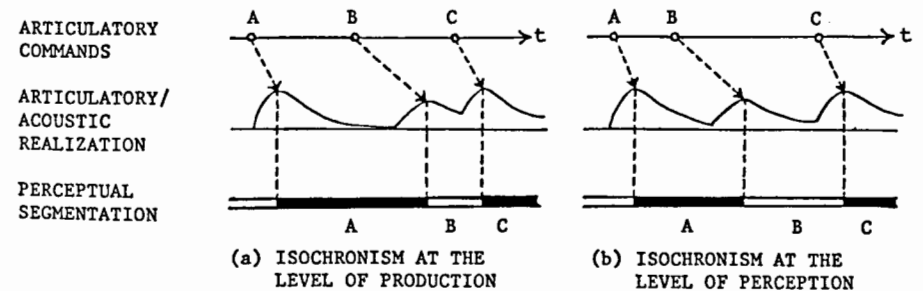


Fig. 7. Two models of the mechanisms underlying the temporal organization of speech units under the open-loop control hypothesis.

with an isochronism, viz., the perceived durations of these segments are kept equal. The results of the present study may be considered as corroborating model (b) as far as the Japanese disyllables are concerned.

Conclusion

Temporal organization of speech segments was investigated using disyllabic Japanese words consisting only of vowels. An acoustic analysis of their formant trajectories has indicated that the onset of the transition to the second vowel in various disyllables is distributed over a relatively wide range. This variation tends to compensate for the differences in the rate of transition due to differences in the articulator(s) involved and the direction of movement. On the other hand, a perceptual analysis of the onset of the second vowel has indicated that the perceptual onset of the second vowel in utterance samples of the same disyllable is concentrated within a relatively narrow range regardless of the particular vowel combination or the order of the vowels in the disyllable. The implication of these findings for the possible mechanisms underlying the temporal organization of speech units was discussed in connection with two models already proposed with regard to these mechanisms.

References

- Fujisaki, H. et al. (1974): "Formulation of the coarticulatory process in the formant frequency domain and its application to automatic recognition of connected vowels," Proc. SCS-74 3, 385-392.
- Fujisaki, H. (1977): "Functional models of articulatory and phonatory dynamics," in Articulatory Modeling and Phonetics, R. Carré, R. Descout, and M. Wajskop (eds.), 127-136, G. A. L. F. Group de la Communication Parlée.
- Kozhevnikov, V. A. and L. A. Chistovich (1965): Speech: Articulation and Perception, Moscow: Nauka.
- Leanderson, R. and B. E. F. Lindblom (1972): "Muscle activation for labial speech gestures," Acta Otolaryng. 73, 362-373.
- Lehiste, I. (1977): "Isochrony reconsidered," Journal of Phonetics 5, 253-263.
- Ohala, J. (1970): "Aspects of the control and production of speech," Working Papers in Phonetics 15, UCLA.

SOME EFFECTS ON INTELLIGIBILITY OF INAPPROPRIATE TEMPORAL
RELATIONS WITHIN SPEECH UNITS

A. W. F. Huggins, Bolt Beranek and Newman Inc, 50 Moulton Street,
Cambridge, Mass 02138, U. S. A.

The purpose of this paper is to make two arguments. The first is that, despite several failures to find such effects, badly disturbed speech timing, such as occurs often in the speech of the deaf for instance, is a sufficient cause for catastrophic loss of intelligibility. If the timing is sufficiently disturbed that the listener cannot identify the pattern of stressed syllables in the sentence -- or, perhaps, its rhythmic pattern -- the sentence will be unintelligible even though virtually all of the phonemes are clearly identifiable in subsequent listening. If the listener perceives a stress/rhythmic pattern that is different from that intended by the speaker, he is "garden-pathed" away from the correct utterance, and is not able to recode the individual phonemes into the words they represent before they fade from auditory short-term memory.

The second argument is that a reason for earlier failures to find strong relationships between timing and intelligibility is that a listener cannot estimate the effect of a particular timing distortion on speech intelligibility if he knows what the sentence says. This fact is already well known. It forms the basis of a popular way of impressing an audience with the fidelity of a speech vocoding system: a demonstration tape is prepared in such a way that the audience already knows what the test sentence is before they hear it as processed by the system whose performance is to be proved. What is not so well known is how easy it is to fall into the trap set by this fact. To be blunt, although I was very aware of the effect, I fell into the trap (Huggins, 1978), and if it can happen to me, it can happen to anyone!

Speech of the Deaf

A major reason for trying to understand speech timing is the need to improve the intelligibility of deaf speakers. Faulty timing has been implicated in poor intelligibility by virtually every major study of deaf speech this century, but this knowledge has not led to the development of effective training methods.

The most frequently cited ways in which the timing of deaf speech differs from normal speech are (1) slower overall rate; (2) more and longer pauses, often inappropriately placed; (3) inadequate differentiation of stressed and unstressed syllables; and (4) excessive lengthening of some segments, especially stops and fricatives (e.g. Nickerson, 1975). Let us consider the foregoing factors in order. Deaf speakers normally take much longer to produce a specified utterance than do normal-hearing speakers. But to the extent that the slower rate is a result of linear stretching of the time scale, slower speech should be more rather than less intelligible. One usually speaks slower (and also more precisely) to someone who has difficulty understanding, such as a child or a foreigner. Furthermore, when recorded speech is instrumentally expanded in time by a factor of four, intelligibility is not affected although the speech becomes tedious to listen to.

Similarly, it would be very surprising if the addition of appropriately placed pauses had a degrading effect on intelligibility. Pauses can be used to mark explicitly the boundaries between groups of syntactically related words. Boundaries so marked need not be inferred from more subtle cues, and the presence of syntactically appropriate pauses should therefore simplify rather than complicate reception. Further, the pauses effectively give the listener additional time to decode the message, and this too lightens rather than increases the processing load (Aaronson et al, 1971).

The occurrence of inappropriate pauses raises a different issue. Inappropriate pauses occur also in normal speech, where they are interpreted as hesitation pauses. These do not appear to interfere with intelligibility. However, listeners are much more sensitive to the presence of inappropriate than appropriate pauses, the threshold for their detection being almost five times smaller (Boomer and Dittmann, 1962). Presumably, then, if inappropriate pauses were interpreted as hesitation pauses in deaf speech also, no damage would result. Problems would arise, however, if the inappropriate pauses were interpreted as appropriate pauses, because this would signal incorrect segmentation of the message. This argument leads to rather a

different view of how timing errors might interfere with intelligibility: they might introduce misleading information about the message which, once accepted, could not be discarded.

There are other aspects of deaf speech which support such a view. Due to difficulties in coordinating different articulators, deaf speakers often produce sounds extraneous to the required sequence, particularly in making and releasing stops and fricatives (Hudgins and Numbers, 1942). If the listener accepts these extraneous sounds as segments, he cannot then go back and delete them. The perceptual apparatus is very good at filling in missing information, but it is very bad at discarding extraneous information unless it occurs as part of a separate auditory "stream" (Bregman and Campbell, 1971). Thus, listeners will swear that they heard a particular segment in a sentence even though it had been totally removed and replaced with an extraneous sound such as a cough (Warren et al, 1969). But the cough cannot be located in the sentence with any accuracy, since it cannot be integrated into a single stream with the speech. When wanted and unwanted segments arrive in a single auditory stream, as they often do in deaf speech, the listener cannot selectively accept the wanted and reject the extraneous segments, even if he had some way of so classifying the segments as they arrived. van Noorden (1975) has shown that two melodies in the same pitch range cannot be identified if they are played by interleaving the notes from the two melodies. The listener cannot decide to listen to alternate notes. On the contrary, he hears only a single sequence. But if one melody is gradually raised in pitch, the two melodies eventually split into two streams, permitting one to be ignored so that the other melody can be recognized.

The listener is not able to discard some of the information after it has been processed, either, and recent models of speech perception offer an explanation. Jarvella (1971) has shown that the accuracy of a listener's verbatim memory for a continuously presented message shows a sharp drop at the preceding clause boundary, as if the need to keep the raw acoustic data available in short-term memory ends when the clausal material is successfully parsed. Thus, any misinterpretations of the

preceding clause that become apparent later cannot easily be corrected, since the verbatim material necessary to the correction has been deleted from short term memory. Furthermore, if the received sequence of segments fails to trigger recognition of a word, the segments fade quite rapidly from auditory short term memory.

When the foregoing arguments are put together with the known importance of correct stress patterns for recognition of words, the poor intelligibility of deaf speech becomes much easier to understand. The pattern of stresses in a word or phrase is of critical importance to its correct recognition. In fact, there is evidence that listeners will discard correctly-heard segmental cues which they cannot reconcile with the perceived stress pattern. English listeners trying to identify English words and phrases, spoken with inappropriate stress patterns by Indian speakers, consistently produced words that matched the incorrect stress patterns, while correct phonemes occurred in enough of the responses to demonstrate that the necessary segmental cues were in fact present (Bansal, 1966). Second, it is known that timing is a vital cue in the perception of stress, outweighing both intensity (loudness) and pitch (Fry, 1958).

Yet it is not clear how much deaf speakers know about stress patterns. For normal listeners, the stress pattern of a word is centrally involved in its memory coding (Brown and McNeill, 1966). It is unlikely that the deaf use a similar coding without being explicitly taught it. Deaf children do not code letters, presented visually in an immediate recall task, in terms of their auditory and articulatory properties, as do normal hearing children and adults (Conrad and Rush, 1965). If the deaf subjects do not use an auditory or articulatory coding scheme for segments, it is very likely that they also use a different coding scheme for stress patterns -- if, indeed, they have a coding scheme for stress patterns at all. Unless the stress pattern of a word is a central part of its representation in memory, the stress pattern is not likely to be reflected in the required pattern of syllable timing when the word is spoken. Yet this pattern of syllable timing is crucial to the intelligibility of the word for hearing listeners.

There are two aspects of incorrect timing that should be distinguished. One type can be traced directly to the difficulty of programming a rapid sequence of articulations. Timing errors become more frequent and more severe as the sentence to be uttered is made more difficult to articulate. The remedy may lie in trying to teach words as integrated motor patterns, and practicing their production first in isolation and then by substituting them in overlearned phrase or sentence frames. This is particularly important in the case of function words, whose fluency in deaf speech is a major determinant of intelligibility (Monson and Leiter, 1975). Timing errors of the foregoing type could be labeled errors of performance, since the deaf speaker is presumably at least partly aware that his production has fallen short of what was intended. The other aspect of incorrect timing is more important, and errors of this type could be labeled errors of intention. Errors of intention occur if the deaf speaker's model of how speech should be timed is different from that of a hearing speaker. In particular, the model may not incorporate the rules for assigning relative stress levels, and for realizing these in timing patterns.

Some evidence supporting the importance for intelligibility of differentiating stressed and unstressed syllables has been reported by Osberger (1978). She produced slight improvements in intelligibility by editing deaf speech waveforms to correct inadequate differentiation of stressed and unstressed syllables. Her method, however, was unable to separate errors of performance from errors of intention, which may account for the smallness of her effects. Also, she reported no attempt to relate the magnitude of the timing corrections made in individual words to the resulting changes in intelligibility.

I have reported elsewhere a preliminary attempt to measure the effects of errors of intention uncontaminated by errors of performance, using synthetic speech (Huggins, 1978). Simple sentences were synthesized in two versions. In one, stress was correctly assigned, and in the other, unstressed syllables were assigned primary stress, and vice versa. Syllables with secondary stress were not affected. Since the same set of synthesis rules were used for stressed as for unstressed

syllables, any errors of performance that were inherent in the synthesis procedure should have affected the normal and mis-stressed versions equally. But when stress was wrongly assigned, word intelligibility fell from 85% to 50%, and the percentage of sentences "substantially understood" fell from 75% to 25%. The results were not uniform across test sentences, in part because the sentences differed in the proportion of syllables carrying primary, secondary, and un-stress, and in part because of some residual errors in phonetic transcription of the test sentences (which may well account for the less than perfect intelligibility of the normally stressed versions). I hope to correct some of these weaknesses in time for the meeting.

Finally, I want to repeat an anecdote from the study. I have tried several times to make a tape demonstrating how unintelligible speech can become when its timing is wrong, but I have never been satisfied with the results. In fact, I began to wonder if what I was trying to show was true. But when I played the latest tape to a colleague, looking for sympathy, he found it totally unintelligible. The difference between us was that I knew what each test sentence said, and therefore knew its stress pattern, whereas he did not. I would never have run the formal experiment but for his unexpected reaction. How many interesting timing effects have been overlooked, or regarded as too slight to be of interest, for similar reasons?

References

- Aaronson, D., N. Markowitz, and H. Shapiro (1971): "Perception and immediate recall of normal and "compressed" auditory sequences," Perception and Psychophysics, 9, 338-344.
- Bansal, R. K. (1966): The intelligibility of Indian English: measurements of the intelligibility of connected speech, and sentence and word material, presented to listeners of different nationalities, Unpublished Ph. D. Thesis, London University.
- Boomer, D. S. and A. T. Dittmann (1962): "Hesitation pauses and juncture pauses in speech," Language and Speech, 5, 215-220.
- Bregman, A. S. and J. Campbell (1971): "Primary auditory stream segregation and perception of order in rapid sequences of tones," J. Experimental Psychology, 89, 244-249.
- Brown, R. and D. McNeill (1966): "The tip of the tongue phenomenon," J. Verbal Learning and Verbal Behavior, 5, 325-337.

- Conrad, R. and M. L. Rush (1965): "Nature of short-term memory encoding by the deaf," J. Speech and Hearing Disorders, 30, 335-343.
- Fry, D. B. (1958): "Experiments on the perception of stress," Language and Speech, 1, 126-152.
- Hudgins, C. V. and F. C. Numbers (1942): "An investigation of intelligibility of speech of the deaf," General Psychology Monograph, 25, 289-392.
- Huggins, A. W. F. (1978): "Speech timing and intelligibility," in J. Requin (ed), Attention and Performance VII, Hillsdale, N.J.: Erlbaum.
- Jarvella, R. (1971): "Syntactic processing of connected speech," J. Verbal Learning and Verbal Behavior, 10, 409-416.
- Monson, R. B. and E. Leiter (1975): "Comparison of intelligibility with duration and pitch control in the speech of deaf children," J. Acoust. Soc. Amer., 57, S69 (A).
- Nickerson, R. S. (1975): "Characteristics of the speech of deaf persons," Volta Review, 77, 342-362.
- Osberger, M. J. (1978): The effect of timing errors on the intelligibility of deaf children's speech. Unpublished doctoral thesis, City University of New York.
- van Noorden, L. P. A. S. (1975): Temporal coherence in the perception of tone sequences. Eindhoven, Netherlands: Technische Hogeschool (Doctoral thesis).
- Warren, R. M., C. J. Obusek, R. M. Farmer, and R. P. Warren (1969): "Auditory sequence: confusions of patterns other than speech or music," Science, 164, 586-587.

SYNTHESIS BY RULE OF SEGMENTAL DURATIONS IN ENGLISH SENTENCES

Dennis H. Klatt, Mass. Inst. of Tech., Cambridge, MA 02139.

In this paper, we are concerned with prediction of the (acoustically defined) durations of phonetic segments in spoken sentences. The durational definitions that have been adopted correspond to the closure for a stop (any burst and aspiration at release are assumed to be a part of the following segment). For fricatives, the duration corresponds to the interval of visible frication noise (or to changes in the voicing source if no frication is visible). For sonorant sequences, the segmental boundary is defined to be the half-way point in the formant transition for that formant having the greatest extent of transition. The definitions represent a convenient largely reproducible measurement procedure, but the physiological and perceptual validity of these boundaries have not been established.

In a review of the factors that influence segmental durations in spoken English sentences (Klatt, 1976a and references cited therein), it was concluded that only some of the systematic durational changes were large enough to be perceptually discriminable. The goal of this paper is to describe these first-order effects by rules.

Input Representation for a Sentence

The durational rule system to be presented is a part of a speech synthesis by rule program (Klatt, 1976b). The phonological component of this program accepts as input an abstract linguistic description of the utterance to be synthesized. The output of the phonological component is a detailed phonetic and prosodic representation of the utterance, including an acoustic duration for each segment. The symbol inventory is shown in Table 1; it includes 52 phonemes, 3 stress markers, 3 types of boundary indicators, and 6 syntactic structure indicators. An example of the use of some of these symbols is provided in Figure 1.

Phonemic Inventory. A traditional phonemic analysis of English is assumed, except that:

- (a) Vowel+/R/ syllables are transcribed with the special vowel nuclei /IR/ ("beer"), /ER/ ("bear"), /AR/ ("bar"), /OR/ ("boar"), and /UR/ ("pure"). Words like "player" and "buyer" should be transcribed with two syllables, i.e. /EY+/RR/ and /AY+/RR/.

(M #F DH AX			#C 1 OW L D	#C M 1 AE N)N	#C S 1 AE T
#F IH N			#F AX	#C R 1 AA K RR)	
Phone	Stress	Dur	Phone	Stress	Dur
SI	0	200	AE	1	165
DH	0	40	DX	0	20
IY	0	85	IH	0	65
OW	1	145	N	0	50
LX	0	65	AX	0	65
D	0	35	R	1	80
M	1	70	AA	1	140
AE	1	225	K	0	50
N	0	60	RR	0	175
S	1	105	SI	0	200

Figure 1. Input representation for "The old man sat in a rocker" and a listing of the output of the phonological component, i.e. the phonetic string, stress feature, and duration predictions in msec.

- (b) The glottal stop [Q], dental flap [DX], glottalized alveolar stop [TQ], and velarized lateral [LX] listed in Table 1 are not really phonemes, but are allophones that are inserted in lexical forms before segmental durations are computed.

Lexical Stress. Each stressed vowel of an utterance must be preceded by a stress symbol (1, 2, or !), where 1 is primary lexical stress (reserved for vowels in open-class content words, only one 1-stress per word). The secondary lexical stress "2" is used in some content words (e.g. the first syllable of "demonstration"), in compounds (e.g. the second syllable of "baseball"), in the strongest syllable of polysyllabic function words (e.g. "until"), and for pronouns (excluding personal pronouns like "his"). Emphatic stress "!" can be assigned to a semantically prominent syllable in a phrase.

Morpheme and Word Boundaries. There is no input symbol to indicate a syllable boundary. The symbol "*" can be used to mark morpheme boundaries. Each word of an utterance to be synthesized must be immediately preceded by a word boundary symbol. The distinction between content and function words is indicated by using "#C" and "#F". Open-class words (nouns, verbs, adjectives and adverbs) are content words. The program will check to see that no function word carries primary stress. A compound such as "apple cart" is indicated in the input representation by replacing the word boundary between "apple" and "cart" by a morpheme boundary and by reducing the lexical stress on the second word "cart" by one.

Table 1. The legal input symbols for synthesis of an utterance. Also given are a basic or inherent duration for each phonetic segment type and a minimum stressed duration in msec.

Vowels		INH DUR	MINDUR			INH DUR	MINDUR
IY	beet	160	50	IH	bit	130	40
EY	ba <u>i</u> t	190	70	EH	b <u>e</u> t	150	60
OW	bo <u>a</u> t	220	70	AH	b <u>u</u> t	140	50
UW	bo <u>o</u> t	210	60	UH	bo <u>o</u> k	160	50
AE	ba <u>a</u> t	230	60	AA	Bo <u>b</u>	240	80
AO	bo <u>u</u> ght	240	80	RR	bi <u>r</u> d	180	60
AY	bi <u>t</u> e	250	90	AW	bo <u>u</u> t	260	100
OY	bo <u>y</u>	280	110	YU	bea <u>u</u> ty	230	100
AX	ab <u>o</u> ut	120	40	IR	be <u>e</u> r	230	100
ER	be <u>a</u> r	270	100	AR	ba <u>r</u>	260	100
OR	bo <u>a</u> r	240	100	UR	po <u>o</u> r	230	100
<u>Sonorant Consonants</u>							
W	w <u>e</u> t	80	60	Y	y <u>e</u> t	80	40
R	r <u>e</u> nt	80	30	L	l <u>e</u> t	80	40
WH	w <u>h</u> ich	70	60	H	h <u>a</u> t	80	20
EL	bo <u>t</u> t <u>l</u> e	160	110	LX	bi <u>l</u> l	90	70
<u>Nasals</u>							
M	m <u>e</u> t	70	60	N	n <u>e</u> t	65	35
NG	si <u>n</u> g	80	50	EM	ke <u>e</u> p' <u>e</u> m	170	110
EN	bu <u>t</u> t <u>e</u> n	170	100				
<u>Fricatives</u>							
F	f <u>i</u> n	120	60	V	v <u>a</u> t	60	40
TH	th <u>i</u> n	110	40	DH	th <u>a</u> t	50	30
S	s <u>a</u> t	125	50	Z	z <u>o</u> o	75	40
SH	sh <u>i</u> n	125	50	ZH	az <u>u</u> re	70	40
<u>Plosives</u>							
P	p <u>e</u> t	85	50	B	b <u>e</u> t	80	50
T	t <u>e</u> n	65	40	D	d <u>e</u> bt	65	40
K	co <u>r</u> e	65	50	G	g <u>o</u> re	65	50
DX	bu <u>t</u> ter	20	20	TQ	at Alan	65	50
Q	Ma <u>o</u> pted	20	20				
<u>Affricates (closure, frication)</u>							
CH	ch <u>i</u> n	70	50	J	g <u>i</u> n	70	50
		60	40			30	20
<u>Stress Symbols</u>							
1	primary lexical stress						
2	secondary lexical stress						
!	emphatic stress						
<u>Word and Morpheme Boundaries</u>							
*	morpheme boundary						
#C	begin content word						
#F	begin function word						
<u>Syntactic Structure</u>							
.	end of declarative utterance						
)?	end of yes/no question						
(M	begin main clause						
,	orthographic comma						
)N	end of noun phrase						
(R	begin relative clause						

Syntactic structure. Syntactic structure symbols are important determiners of sentence stress, rhythm, and intonation. Syntactic structure symbols appear just before the word boundary symbol. Only one syntactic marker can appear at a given sentence position. The strongest syntactic boundary symbol is always used (the stronger symbols appear higher in the list in Table 1).

An utterance must end with either a period "." signalling a final fall in intonation, or a question mark ")" signalling the intonation pattern appropriate for yes-no questions. Each clause must be preceded by either "(M" to indicate the beginning of a main clause, or "(R" to indicate the beginning of a relative clause. If clauses are conjoined, a syntactic symbol is placed just before the conjunction. If a comma could be placed in the orthographic rendition of the desired utterance, then the syntactic comma symbol "," should be inserted. Syntactic commas are treated as full clause boundaries in the rules; they are used to break up larger units into chunks in order to facilitate perceptual processing. The end of a noun phrase is indicated by ")N". Segments in the syllable prior to a syntactic boundary are lengthened. Based on the results of Carlson, Granstrom, and Klatt (1979), an exception is suggested in that any)N following a noun phrase that contains only one primary-stressed content word should be erased. The NP + VP is then spoken as a single phonological phrase with no internal phrase-final lengthening.

Rules

The representation for a sentence discussed above serves as input to the phonological component of the synthesis-by-rule program. The form of the output from the phonological rules is shown at the bottom in Figure 1. The abstract string of symbols has been converted to a string of phonetic segments, with each segment being assigned a stress feature and duration in msec. Before presenting details of the duration algorithm, we summarize some of the rules that must be executed prior to duration prediction.

Stress Rules. The phonological component assigns a feature Stress (value = 0 or 1) to each phonetic segment in the output string. The default value is 0 (unstressed). Vowels preceded by a 1 or 2-stress in the input are assigned a value of 1. Consonants

preceding a stressed vowel are also assigned a value of 1 if they are in the same morpheme and if they form an acceptable word-initial consonant cluster. Segmental stress is used in rules that determine segmental duration, fundamental frequency, plosive aspiration duration, and formant target undershoot.

Rules of Segmental Phonology. There are presently very few phonological rules of a segmental nature in the program. A number of rules that are sometimes attributed by linguists to the phonological component (e.g. palatalization) are realized in the phonetic component because they involve graded phenomena (e.g. the [S] of "fish soup" is partially palatalized, but not identical to [SH:]. The segmental (within-word and across-word-boundary) phonological rules that are described below are extremely important. They are not "sloppy speech" rules, but rather rules that aid the listener in hypothesizing the locations of word and phrase boundaries. For example, the second rule ensures that a word-final /T/ is not perceived as a part of the next word by inserting simultaneous glottalization to attenuate any release burst. Rules are expressed in a feature-based notation that is compiled into Fortran code for computer simulation of the phonological component (Klatt, 1976b). Rules 1 and 2 below are stated in this way, while the others are expressed in ordinary English.

1. [L] --> [LX]/(+VOWEL)...(-STRESS)
Substitute a postvocalic velarized allophone [LX] for [L] if the [L] is preceded by a vowel and followed by anything except a stressed vowel in the same word.
2. ([T] or [D]) --> [DX]/(+SONOR -NASAL)...(-STRESS +VOWEL)
Replace [T] or [D] by the alveolar flap [DX] within words and across words boundaries (but not across phrase and clause boundaries) if the plosive is followed by a non-primary-stressed vowel and preceded by a nonnasal sonorant. Examples: "butter", "ladder", "sat about".
3. A word-final [T] preceded by a sonorant is replaced by the glottalized dental stop TQ (i.e. has a glottal release rather than a t-burst) if the next word starts with a stressed sonorant (unless there is a clause boundary between the words, in which case the [T] is released into a pause). Examples: "that one", "Mat ran".
4. A voiceless plosive is not released if the next phonetic segment is another voiceless plosive within the same clause.
5. A glottal stop [Q] is inserted before a word-initial stressed vowel if the preceding segment is syllabic (and not a determiner), or if the preceding segment is a voiced nonplosive and there is an intervening phrase boundary. Example: "Liz eats".

6. Unstressed [OR] is replaced by syllabic [RR], as in "for him" or "forget". (There are many rules of this type.)

Duration Rules. Each segment is assigned a duration by a set of rules presented in detail below. The rules are intended to match observed durations for a single speaker (DHK) reading paragraph-length materials. The rules operate within the framework of a model of durational behavior which states that (1) each rule tries to effect a percentage increase or decrease in the duration of the segment, but (2) segments cannot be compressed shorter than a certain minimum duration (Klatt, 1976a). The model is summarized by the formula:

$$DUR = ((INH DUR - MINDUR) * PRCNT) / 100 + MINDUR \quad (1)$$

where INHDUR is the inherent duration of a segment in msec, MINDUR is the minimum duration of a segment if stressed, and PRCNT is the percentage shortening determined by applying rules 1 to 10 below. The program begins by obtaining values for INHDUR and MINDUR for the current segment from Table 1, and by setting PRCNT to 100. The inherent duration has no special status other than as a starting point for rule application; it is roughly the duration to be expected in nonsense CVCs spoken in the carrier phrase "Say bVb again" or "Say CaC again". The following ten rules are then applied, where each rule modifies the PRCNT value obtained from the previous applicable rules according to the equation:

$$PRCNT = (PRCNT * PRCNT1) / 100 \quad (2)$$

The duration of the segment is then computed by inserting the final value for PRCNT into Equation 1 and, finally, Rule 11 is applied.

1. PAUSE INSERTION RULE: Insert a 200 msec pause before each sentence-internal main clause and at boundaries delimited by a syntactic comma, but not before relative clauses. The "(R" symbol functions like a "N" in the duration rules.
2. CLAUSE-FINAL LENGTHENING: The vowel or syllabic consonant in the syllable just before a pause is lengthened by PRCNT1=140. Any consonants between this vowel and the pause are also lengthened by PRCNT1=140.
3. NON-PHRASE-FINAL SHORTENING: Syllabic segments (vowels and syllabic consonants) are shortened by PRCNT1=60 if not in a phrase-final syllable. A phrase-final postvocalic liquid or nasal is lengthened by PRCNT1=140.
4. NON-WORD-FINAL SHORTENING: Syllabic segments are shortened by PRCNT1=85 if not in a word-final syllable.
5. POLYSYLLABIC SHORTENING: Syllabic segments in a polysyllabic word are shortened by PRCNT1=80.

6. NON-INITIAL-CONSONANT SHORTENING: Consonants in non-word-initial position are shortened by $PRCNT1=85$.
7. UNSTRESSED SHORTENING: Unstressed segments are half again more compressible than stressed segments (i.e. set $MINDUR=MINDUR/2$). Then both unstressed and 2-stressed segments are shortened by a factor $PRCNT1$ that is tabulated below for each type of segment. The result is that segments assigned secondary stress are shortened relative to 1-stress, but not as much as unstressed segments.

Context	PRCNT1 for Unstr. and 2-stress
syllabic (word-medial syll)	50
syllabic (others)	70
prevocalic liquid or glide	10
all others	70

8. LENGTHENING FOR EMPHASIS: An emphasized vowel is lengthened by $PRCNT1=140$ percent.
9. POSTVOCALIC CONTEXT OF VOWELS: The influence of a postvocalic consonant or sonorant-stop cluster on the duration of a vowel is given below. (Cs must be in the same morpheme as the V and must have the feature unstressed.) In a postvocalic sonorant-obstruent cluster, the obstruent determines the effect on the vowel and on the sonorant.

Context	PRCNT1
open syllable, word-final	120
before a voiced fricative	160
before a voiced plosive	120
before an unstr. nasal	85
before a voiceless plosive	70
all others	100

The effects are greatest at phrase and clause boundaries: if non-phrase-final, change $PRCNT1$ to be $70 + 0.3*PRCNT1$

10. SHORTENING IN CLUSTERS: Segments are shortened in consonant-consonant sequences (disregarding word boundaries, but not across phrase boundaries), and segments are also modified in duration in vowel-vowel sequences.

Context	PRCNT1
vowel followed by a vowel	120
vowel preceded by a vowel	70
consonant surrounded by consonants	50
consonant preceded by a consonant	70
consonant followed by a consonant	70

11. LENGTHENING DUE TO PLOSIVE ASPIRATION: A 1-stressed or 2-stressed vowel or sonorant preceded by an aspirated plosive is lengthened by 25 msec.

When the rules are applied to the /RR/ of "rocker" in Figure 1, the second rule sets $PRCNT$ to 140, the fifth rule reduces $PRCNT$ to 112, the seventh rule reduces $MINDUR$ to 30 msec and $PRCNT$ to 78.4, and the ninth rule increases $PRCNT$ to 94. Then $INHUR$, $MINDUR$, and $PRCNT$ are inserted in Equation 1 and the resulting duration is rounded up to the nearest 5 msec to obtain the value of 175 msec.

The resulting durations are determined in part by a variable that controls the nominal speaking rate $SPRATE$ which can be set to any number between 60 and 300 words per minute. The default value is 180 words per minute. At rates slower than 150 wpm, a short pause is inserted between a content word and a following function word. (At a normal speaking rate, brief pauses are inserted only at the ends of clauses.) Individual segments are lengthened or shortened slightly depending on speaking rate, but most of the rate change is realized by manipulating pause durations.

Evaluation

The rules constitute only a first-order approximation to many of the durational phenomena seen in sentences (e.g. consonant interactions in clusters) and the rules completely ignore other factors. Nevertheless, as a first approximation, the rules capture a good deal of the systematic variation in segmental durations for speaker DHK. When compared with spectrograms of new paragraphs read by this speaker, the rule system produces segmental durations that differ from measured durations by a standard deviation of 17 msec (excluding the prediction of pause durations), and the rules account for 84 percent of the observed total variance in segmental durations. Seventeen msec is generally less than the just-noticeable difference for a single change to segmental duration in sentence materials (Klatt, 1976a).

A perceptual evaluation of the performance of the rule system is discussed by Carlson, Granstrom and Klatt (1979). The perceptual results are encouraging in that both naturalness and intelligibility ratings of sentences synthesized by these rules are very similar to ratings of the same sentences synthesized using durations obtained from a natural recording.

References

- Carlson, R., Granstrom, B., and Klatt, D.H. (1979), "Some Notes on the Perception of Temporal Patterns in Speech", 9th International Congress of Phonetic Sciences, Copenhagen.
- Klatt, D.H. (1976a), "Linguistic Uses of Segmental Duration in English: Acoustic and Perceptual Evidence", J. Acoust. Soc. Am. 59, 1208-1221.
- Klatt, D.H. (1976b), "Structure of a Phonological Rule Component for a Speech Synthesis by Rule Program", IEEE Trans. Acoustics, Speech, and Signal Processing ASSP-24, 391-398.

COMPLEX CONTROL OF SIMPLE DECISIONS IN THE PERCEPTION OF VOWEL LENGTH

Sieb G. Nootboom, Institute for Perception Research, Eindhoven Netherlands

Introduction

Measurable vowel durations in connected speech are influenced by many different factors. Well-known examples are the effects on the durations of stressed vowels of (1) the postvocalic consonants, (2) the position of the syllable in the word, (3) the syntactic position of the word in the sentence, (4) the presence or absence of a speech pause following the syllable to which the vowel belongs, (5) overall speech rate. Despite the wide variability in acoustic durations due to the combined effects of these and other factors, in many languages vowel durations are contrastive cues to vowel phoneme perception. The fact that the shortest durations of phonemically long vowels can be considerably shorter than the longest durations of phonemically short vowels does apparently not prevent listeners from making correct decisions as to perceived phonemic vowel length. Let us examine how a decision strategy might be organized in order to accomplish this.

A hypothesized decision strategy

We assume, in terms of signal detection theory, that each acoustic vowel duration is represented on an internal stimulus strength axis with an uncertainty that is equal for all vowel durations, at least within the limited range of durations in which we are interested. This uncertainty, due the effects of sensation noise, gives rise to a Gaussian distribution of stimulus strength values for repetitions of each particular vowel duration. We further assume that identification of vowel length is so organized that each individual stimulus strength value X , derived from a single vowel duration, is compared to an internal criterion C on the stimulus strength axis. All X lower than C are identified as short vowel phonemes, all X higher than C are identified as long vowel phonemes. The criterion C is assumed here to be noiseless so that all uncertainty in phonemic decisions has to be caused by the effects of sensation noise. Due to the Gaussian form of the stimulus strength distributions, the probability distribution of short vowel decisions over a set of acoustic vowel durations, and of course the complementary probability distribution of long vowel decisions,

will have the form of a cumulative normal distribution with a mean μ (the phoneme boundary) reflecting the position of the internal criterion C on the stimulus strength axis, and a standard deviation σ reflecting the effect of sensation noise.

Although the internal criterion C is supposed to be noiseless, this does not imply that it has always the same position on the stimulus strength axis. We assume that the listener can move the internal criterion up and down the stimulus strength axis, adjusting its position in order to optimize the chance of correct recognition. Imagine that a listener perceives a vowel segment and is not sure whether the phoneme intended by the speaker was a long or a short vowel. He may then take into account that the vowel concerned is clearly stressed, is followed by a voiceless plosive, is in the final syllable of a word which is immediately followed by a major syntactic boundary, not accompanied by a speech pause, and that the speech rate is slightly faster than normal. Our listener knows from experience that in these conditions a short vowel would have had a stimulus strength value A and a long vowel a stimulus strength value B . He therefore places his internal criterion C in the middle between A and B , in this way optimizing his chance of a correct decision on perceived vowel length.

Of course, such a decision strategy implies that listeners have an extensive and detailed knowledge of the temporal regularities of speech and are able to apply this knowledge very rapidly, so rapidly in fact that they are not aware of doing so. They are even unaware of having this knowledge. We cannot, therefore, test the proposed theory by asking listeners what they do. We need another kind of test. Let us examine a few specific hypotheses that may be derived from the theory and see whether they are corroborated by experimental data.

The effect of sensation noise

If our theory is correct we could measure the accuracy of auditory representation in a vowel length identification task, by determining the σ of the cumulative normal distribution. Of course, in psychoacoustics the accuracy of auditory representation is generally expressed in terms of a differential threshold measured in a binary forced choice comparison task, involving two stimuli per decision. Our theory predicts that the accuracy is equal in both tasks, because there is no reason to suppose that the effect of

sensation noise would be different. In testing this hypothesis, however, we must take into account that in a binary forced choice task involving two stimuli the stimulus separation needed to obtain a given level of performance, for example the 75 % level, is $\sqrt{2}$ greater than the stimulus separation needed to obtain the same level of performance in a similar task involving only one stimulus per decision (Green and Swets, 1966, 68). Because the σ of a cumulative normal distribution is almost $\sqrt{2}$ times the 75 % differential threshold, we may compare the σ measured in a binary forced choice identification task directly to 75 % differential thresholds measured in a comparison task.

In a number of identification tests on the effect of acoustic vowel duration on the distinction between Dutch short /a/ and long /a:/ in different speech contexts, ranging from isolated vowels to vowels embedded in full sentences, we have found σ 's averaged over groups of at least 10 listeners for each context condition, between 10 ms (for vowels in isolation) and 3 ms (for one particular full-sentence condition). In most conditions σ 's were in the order of 6 ms. Phoneme boundaries ranged from 75 to 100 ms. These σ 's are within the range of differential thresholds of sound duration reported in the literature for sounds with approximately the same durations (Lehiste, 1970; Nootboom and Doodeman, 1978). There is an unpredicted and clear tendency for the σ 's to decrease when more speech context becomes available to the listeners. If we stick to our assumption that, within each particular context, the internal criterion C is noiseless, this would mean that the effect of sensation noise is context dependent. If one would prefer to assume that the effect of sensation noise is independent of speech context, one would have to abandon the idea of a noiseless criterion, and assume that the internal criterion shows less uncertainty with increasing embeddedness of the vowel segment.

Moving the internal criterion up and down

Let us now see what happens to the phoneme boundary, being the measurable reflection of the assumed internal criterion C in the decision strategy, when we change the speech context. The proposed strategy implies that, when the speech context changes in such a way that the expected durations of short and long vowel phonemes change, the internal criterion will move up and down accordingly. This prediction has been tested for changes in speech context re-

lated to (1) the postvocalic consonant, (2) the position of the syllable in the word, (3) the syntactic position of the word in the sentence, (4) the presence or absence of a speech pause after the syllable, (5) the overall speech rate. The experimental design was very similar in all cases and has been described in detail elsewhere (Nootboom and Doodeman, 1978). All experiments were limited to the Dutch /a/ - /a:/ opposition. It should be noted that in natural speech these two vowels are distinguished not only by their relative durations but also by their spectral properties, which were kept constant in the experiments. All experiments involved at least 10 subjects. Let us briefly review the results.

- The postvocalic consonant

Vowel segments followed by a speech pause have generally a greater duration than those followed by a consonant. The amount of shortening caused by the postvocalic consonant depends on the nature of the consonant. For example, plosives tend to shorten the preceding vowel more than do fricatives. This is valid for both short and long vowel phonemes. Consequently the optimal position of the internal criterion C will be at a lower stimulus strength value for vowels followed by a fricative than for vowels in isolation, and at a still lower value for vowels followed by a voiceless plosive. We therefore can predict that the phoneme boundary between /a/ and /a:/ in isolation lies at a greater duration than the same phoneme boundary measured before /s/, which again lies at a greater duration than the one measured before /t/. This prediction is corroborated by the data. We find the following phoneme boundaries, estimated from probability distributions in a vowel length identification test by fitting cumulative normal distributions (sd stands for the standard deviation over the subjects, and is not to be confused with the earlier discussed σ):

In isolation 100 ms (sd 8.4 ms)

Before /s/ 97 ms (sd 6.7 ms)

Before /t/ 91 ms (sd 6.9 ms)

- The position in the word

Both short and long Dutch vowels bearing lexical stress tend to become shorter with increasing number of unstressed syllables following in the word (Nootboom, 1973). Thus we predict that the phoneme boundary will shift towards shorter durations when more

unstressed syllables are added to the word. This has been tested with nonsense words in isolation, in which the first, stressed, syllable contained the test vowel segment, followed by intervocalic /t/, and the number of unstressed syllables was 0, 1, 2 or 3. Phoneme boundaries and standard deviations over the subjects were:

- 0 unstr. syll. 91 ms (sd 6.9 ms)
- 1 unstr. syll. 88 ms (sd 5.8 ms)
- 2 unstr. syll. 85 ms (sd 5.5 ms)
- 3 unstr. syll. 83 ms (sd 4.3 ms)

Differences between the last three conditions might have been induced by perceived changes in the second syllable.

- The syntactic position of the word

Durations of Dutch short and long vowels in monosyllabic words vary systematically with syntactic position of the word, notably with the type of syntactic boundary immediately following the word. In one experiment we measured phoneme boundaries between /a/ and /a:/ in a monosyllable /tVk/ embedded in 5 different test utterances, each with a different syntactic structure. These test utterances were obtained from sentences spoken with the long vowel /a:/ in the test segment slot, and had normal rhythm and intonation. This original vowel /a:/ had durations ranging from 150 ms in one utterance to 190 ms in another. In each spoken sentence the original vowel segment was excised and replaced by one of a set of test segments differing in acoustic duration. Phoneme boundaries were assessed in each of the 5 test utterances for each of 12 subjects. They ranged from 76 ms, for the test utterance with an original /a:/-duration of 150 ms, to 100 ms, for the test utterance with an original /a:/-duration of 190 ms. Calculating the product moment correlation of the original /a:/-durations in each of the 5 test utterances with all 12 phoneme boundaries in each of these utterances gave $r = 0.83$ ($p < 0.001$). Apparently the /a:/-durations as originally spoken in these test utterances are to a fair extent controlled by the same factors as the phoneme boundaries in the identification test. These factors are either to be looked for in the syntactic structures of the sentences, as intended by the speaker and perceived by the listeners, or, more probably, in the prosodic structures of the sentence realizations, which, of course, are partly determined by the syntactic structures.

- The prepausal position of the syllable

Durations of Dutch short and long vowels in monosyllabic words are considerably longer when the word is in prepausal position than when it is not, other things being equal. Thus we may predict that the phoneme boundary in an embedded monosyllable will shift towards a greater duration when we insert a speech pause immediately after the monosyllable. This has been tested by inserting acoustic silent intervals with durations of 0, 100, 200, and 800 ms immediately after the test syllable /tVk/ embedded in a test utterance, and measuring the phoneme boundary in each of these conditions. In addition, the probability of speech pause perception has been measured independently. It was found that the phoneme boundary increased from 79 ms for a silent interval of 0 ms, to 94 ms for a silent interval of 800 ms, and that the phoneme boundaries in all test conditions were accurately predicted by

$$pb = 79 + 15P_{spp}$$

in which pb is the phoneme boundary in ms, and P_{spp} is the probability of speech pause perception. The probability distribution of speech pause perception over the durations of silent intervals follows an exponential function with a time constant of 200 ms. One possible interpretation of these results is that the listeners employed in this experiment two discrete internal criteria for vowel length identification, one for the prepausal and one for the non-prepausal condition. The gradual increase of the mean phoneme boundary value could be entirely due to the gradual increase in the probability of speech pause perception.

- The effect of speech rate

The effect of speech rate was assessed in a listening experiment employing a computer-controlled channel vocoder in order to vary overall speech rate of a test utterance. In this test utterance the syllable /tVk/ was in final position and the speech rate of all speech material preceding the test vowel segment was 0.67, 1, or 1.5 times normal. Phoneme boundaries were 81, 95 and 102 ms respectively, showing a partial adjustment to speech rate.

Conclusions

The proposed decision strategy for the disambiguation of vowel length is confirmed in all experimental tests. The effect of sensation noise does not conflict with what would be predicted from

differential thresholds for sound duration, although there is an unexpected tendency towards higher accuracy with increasing amount of embeddedness of the test vowel segment. The effect of speech context on the phoneme boundaries is found to be in the predicted direction in all 5 types of contextual differences which were investigated.

The shifts of the phoneme boundary may seem small, ranging from a few ms to about 25 ms. However, they are generally in the same order of magnitude as contextual effects on the durations of short vowels. Analogous to the tendency towards higher within-subject accuracy with increasing amount of embeddedness, we find less inter-subject variation with increasing amount of embeddedness, suggesting that the position of the internal criterion C becomes more and more constrained by inter-personal factors when more speech context becomes available to the listener.

The results support our hypothesis that listeners, whether they know it or not, have an extensive and detailed knowledge of the temporal regularities of speech and actually apply this knowledge rapidly and unknowingly in optimizing their chance of correct decisions on perceived vowel length. This strategy for disambiguation of vowel length may seem an extremely complex and even cumbersome machinery for the communication of a very simple binary contrast. However, given the complexity of the temporal organization of speech, decision strategies in perception have to be complex in order to be efficient.

References

- Green, D.M. and J.A. Swets (1966): Signal detection theory and psychophysics, New York: Wiley.
- Lehiste, I. (1970): Suprasegmentals, Cambridge, Massachusetts, and London, England: M.I.T. Press.
- Nooteboom, S.G. (1972): "The interaction of some intra-syllable and extra-syllable factors on syllable nucleus durations", Institute for Perception Research Annual Progress Report 7, 30-39.
- Nooteboom, S.G. (1973): "The perceptual reality of some prosodic durations", JPh 1, 25-45.
- Nooteboom, S.G. and G.J.N. Doodeman (1978): "Perception of vowel length in spoken sentences", submitted for publication.

PREDICTING SEGMENT DURATIONS IN TERMS OF A GESTURE THEORY
OF SPEECH PRODUCTION

S.E.G. Öhman, S. Zetterlund, L. Nordstrand and O. Engstrand,
Dept. of Linguistics, Uppsala University, Sweden

Theory

We take the basic object of phonetic investigation to be the concrete sound gestalt produced by a speaker in a speech act, i.e. the spoken sentence. And we take the basic problem of our research to be that of explaining the physical structure of the sentence considered, not merely as a complex sound, but as a complex sound used as a vehicle for linguistic communication, or, briefly speaking, we take the problem to be that of explaining the phonetic structure of the sentence.

We explain the sentence phonetically by giving an explicit account of its linguistically functional, physical components, (atomic and compound), and of the methods by which these components are made to form a whole.

An oscillographic or spectrographic record of a spoken sentence does not by itself explain the phonetic structure of the sentence. In such a record, both the (primary) acoustic effects intended by the speaker to form the linguistically functional (atomic or compound) parts of the sentence, and the (secondary) acoustic traces of the speaker's efforts to bring the intended acoustic effects about, will be visible. As a first step in the phonetic explanation of a sentence we therefore try, on the basis of experiment, to distinguish the former acoustic effects from the latter, to define the intended (primary) effects in acoustic terms, and to explain with reference to the physiological properties of the organs of sound production and perception, why the speaker chooses to bring the intended acoustic effects about in just the way he does. In particular, we require detailed explanations of why the (secondary) acoustic traces of these efforts have the acoustic properties that they have.

As an example of this, consider a phonetic part of a sentence that has the form of a voiceless stop consonant such as [t]. It may be assumed that the intended acoustic effect in this case (the [t] per se) is the brief burst of friction noise. The formant transitions that follow this burst (into a following vowel), the

voiceless time interval that precedes it, and the formant transitions (from a preceding vowel) that precede this voiceless interval, are all to be regarded as secondary acoustic traces of the speaker's efforts to bring the burst about. These traces may be explained by showing that a burst of the type in question can only (or at least, most easily) be produced by building up air pressure behind an oral closure at a certain place and by then quickly releasing this pressure in the familiar manner. (A detailed analysis would of course have to make quantitative predictions on the basis of an explicit production and perception model.)

We evaluate an assumption about what does and what does not constitute the intended (primary) acoustic effects, in the total complex sound of a sentence, (1) on the basis of the possibility of explaining convincingly the detailed physical structure of the sentence (including secondary effects), given the physiological constraints on the production and perception mechanisms, and given that the speaker's goal is that of bringing about the assumed primary effects, and (2) on the basis of the depth of insight that the assumed phonetic structure gives us as regards the semantic function of the sentence in the speech act where it was used.

We do not assume that the phonetic structure of a sentence is segmental, nor that it is linear. Experience indicates, on the contrary, that the following picture is better justified:

In any language one operates with a finite inventory of types of atomic acoustic effects. We write E_1, \dots, E_n to denote these effect types for a given language (with n such types). Moreover, we use lower case letters (such as e) to denote intended (atomic) acoustic effects, and we write

$$(1) \quad e \in E_1 \quad \text{or} \quad E_1(e)$$

to say that the atomic acoustic effect e is of type E_1 .

Atomic acoustic effects may be combined to form larger acoustic units in either of two ways called coarticulation and sequencing. We write

$$(2) \quad e_1 + e_2$$

to indicate that the two atomic acoustic effects e_1 and e_2 are coarticulated, which means that they come about (are brought about) simultaneously, or, more accurately, that there is a point in time at which both these effects are heard. And we write

$$(3) \quad e_1 \bullet e_2$$

to indicate that the two atomic acoustic effects e_1 and e_2 are sequenced, which means that e_1 comes about, whereupon e_2 comes about as soon as can be done. It should be noted that, for all e_i, e_j and e_k the following equalities hold:

$$(4) \quad e_i + e_j = e_j + e_i$$

$$(5) \quad (e_i + e_j) + e_k = e_i + (e_j + e_k)$$

$$(6) \quad (e_i \bullet e_j) \bullet e_k = e_i \bullet (e_j \bullet e_k)$$

$$(7) \quad e_i + e_i = e_i \bullet e_i = e_i$$

Every language will have special rules according to which certain atomic acoustic effects (specific to that language) can be coarticulated and sequenced. These rules will also allow coarticulation and sequencing of nonatomic (compound) acoustic effects (differently in different languages). If e_1 and/or e_2 are compound effects, the expression $e_1 \bullet e_2$ denotes the sequence in which e_2 develops immediately after the last effect of e_1 has emerged; and $e_1 + e_2$ denotes an effect in which e_1 and e_2 develop simultaneously in such a way that the last effects of e_1 and e_2 coincide in time.

In most languages, we should expect to encounter compound acoustic effects of both of the following forms

$$(8) \quad (e_2 + e_5) \bullet e_3$$

$$(9) \quad e_2 + (e_5 \bullet e_3)$$

Here (8) is to be read: the compound effect in which e_2 is coarticulated with e_5 , immediately followed (as a whole) by e_3 . And (9) is to be read: the compound effect in which e_2 is coarticulated with a compound effect in which the effect e_5 is immediately followed by e_3 .

The linearity hypothesis, which we reject, excludes compound effects of the form (9) above.

The acoustic effects are related to articulation as follows. When the speaker says his sentence he knows what acoustic effects he intends to bring about and how they are to be arranged in terms of coarticulation and sequencing. In order to bring these effects about, he makes audible gestures with his organs of speech production, one gesture for each acoustic effect intended, whether atomic

or compound. I.e., the gestures can also be regarded as atomic or compound.

The gestures will be timed and executed in such a manner that (1) the intended acoustic effects come about and (2) no intended acoustic effects are destroyed by the bringing about of other effects.

We hypothesize that in a very considerable number of cases the segmental structure of sentences visible in oscillograms and sound spectrograms and, in particular, the temporal durations of these acoustic segments, can be explained as secondary effects due to the speaker's efforts to bring about the linguistically functional, primary acoustic effects.

The reasoning behind this hypothesis is, among others, this: The linguistically functional, intended acoustic effects are not, in general, required to have any particular duration. They are felt to be complete as soon as they are heard to emerge. A complex acoustic effect in which several atomic effects are coarticulated may, however, require for its execution a compound gesture one part of which is slower than all the others. If several of these gestures are started at about the same time, some of them may be completed earlier than the others in the sense that the effects that they aim at bringing about emerge before the others. To coarticulate all the effects, i.e. make them audible at the same time, the effects that emerge early will have to be maintained for some time while waiting for the remaining effects to materialize. Thus, acoustic segments with quasi-stationary qualities will arise not as a final end of the phonetic action but as a secondary consequence of the effort to reach a certain final end (the simultaneous sounding of the effects in question).

As an example of an alleged phonological contrast that seems eliminable on this paradigm we offer the Swedish contrast [vi:la] [vi:l:a] (rest, house) which we analyze as

v (stress + i) • l • a

v (stress + (i • l)) • l • a

where the stress effect which it takes a relatively long time to produce must be coarticulated with the vowel [i], (thus making the quickly producible [i] long) in the first case, whereas the stress effect is coarticulated with [i • l] in the second case (thus making the [i] long).

Among the acoustic effect types of most languages there will be certain (relative) pitch levels or compounds (sequences) of such levels. These pitch levels will in general be coarticulated with acoustic effects with the feature [+voice], especially vowels. We therefore expect that vowel duration will be strongly dependent on intonation in most languages.

In what follows some experimental data that have been collected to test this theory will be summarized.

Data

In two experiments (Zetterlund et al. 1978, Engstrand et al. 1978) we used a computer system (ILS) for manipulating prosodic parameters in natural speech to show that listeners consistently tend to overlook systematic durational variations in their identification of certain noun phrases such as compounds and lexicalized phrases. In an identification experiment we presented our informants with various synthesized versions of certain utterances (see Zetterlund et al. 1978, Engstrand et al. 1978) systematically changing fundamental frequency, vowel and consonant durations, and intensity. The responses were consistent to almost one hundred percent: The critical parameter for the listeners' identification of these utterances was F0. Although the acoustic analysis displays great variations in the duration and intensity parameters, our subjects apparently paid no attention to these potential cues in the presence of F0.

On the basis of the theory sketched earlier in this paper, we expect that a large F0 movement between two critical values would tend to space these points further apart in time than a small (or no) F0 change. To test this hypothesis we have in one experiment looked at words involving the Swedish word accent opposition. In bisyllabic accent 2 words in focus position F0 has to be low at the end of the first (stressed) vowel. The second vowel carries the sentence accent which is physically signaled as a high F0 at the beginning of the vowel. Consequently, most of the F0 rise has to take place during the intervening consonant occlusion. The corresponding accent 1 words do not display this upward shift but are characterized by a more or less level F0 contour during the consonant. The interesting thing about this is that the consonant in the accent 2 words seems to be significantly longer than the consonants in the corresponding accent 1 words.

It is known that F0 variations are accompanied by considerable vertical movements of the entire larynx box. Combined electro-myographic and larynx movement data that we have collected show that these vertical movements have definite muscular correlates, namely activity in the geniohyoid and sternohyoid muscles for upward and downward movement, respectively. Although the way the vertical movements mechanically affect the tension of the vocal folds is very much open to question, we can state that there is a very high positive correlation between larynx height and F0, and that F0 control involves a delicate coordination between small intrinsic musculature and larger supra- and infrahyoidal muscle masses. And, further, considering the comparatively large mass of the larynx box, it seems rather plausible to assume that its mechanical inertia in combination with the coordinative demands on the muscles should impose restrictions on the velocity with which it can conveniently and accurately be moved.

A further example is this: In an accent 1 word F0 is phonologically required to be low at the beginning of the stressed vowel. If the accent 1 word is preceded in the sentence by a relatively high F0, a downward movement of F0 is observed at the beginning of the accent 1 word sometimes extending into its stressed vowel. The duration of the consonant preceding the stressed vowel is found experimentally strongly to depend on the extent of the pitch drop through the beginning of the accent 1 word.

Finally, in order further to test the theory we have looked at the dispersion of the F0-values at various critical points and found that the standard deviations generally are extremely small, e.g. 2.6 Hz at the last peak of the hu segment in vita huset. We have several more examples of this kind.

Obviously, the most important task must be to establish the critical F0-values at different points in time with greater certainty. The perceptual tolerances that listeners have to deviations in the time-frequency domain should be investigated. We would also like to know what significance the exact shape of the F0 contour between the critical points have to a listener. As a matter of fact, the question whether the entire F0 contour or only some fixed values at certain line-up points relative to supra-glottal articulations are the intended and, therefore, phonologically crucial effects produced by the speaker is not yet completely

answered. Pilot experiments encourage us to believe that the latter hypothesis will prove to be true. This would mean, then, that a speaker is given a relatively large amount of freedom to choose ad hoc strategies for passing through the chain of successive critical F0 points. If this is true, a reasonable assumption is that the way transitions are brought about is adapted to fit some anatomical constraints on the larynx. Looking at our data we observe that the slopes of F0 rises and falls are characterized by constancy rather than variation.

References

- Engstrand, O., L. Nordstrand and S. Zetterlund: Experiments on the perceptual evaluation of prosodic parameters in compounds and lexicalized phrases. Paper given at The Phonetics Symposium held at the Department of Speech Communication, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, November 9-10, 1978.
- Zetterlund, S., L. Nordstrand and O. Engstrand: An experiment on the perceptual evaluation of prosodic parameters for word boundary decision in Swedish. Paper given at The Symposium on the Prosody of the Nordic Languages, Phonetics Laboratory, Department of General Linguistics, Lund University, June 14-16, 1978.
- Öhman, S.: Aktuell svensk forskning i fonetik. Tionde sammankomsten för svenskans beskrivning, Uppsala, april 1977. In: S. Eliasson, B. Loman, B. Sigurd, U. Telemann and S. Öhman: Svenskan i modern belysning. Fem översikter från Tionde sammankomsten för svenskans beskrivning (Ord och stil. Språkvårdssamfundets skrifter 9.) Lund: Studentlitteratur.

MOTOR CONTROL OF SPEECH GESTURES

Summary of Moderator's Introduction

James Lubker, University of Stockholm

Speech production theory is currently faced with several closely related and quite crucial issues which are well illustrated by the papers in this Symposium on Motor Control of Speech Gestures.

Perhaps central to these issues is the growing impatience among many phoneticians with what they see as a constraint to bend or adapt physiological/mechanical "fact" from motor control research to fit abstract linguistic constructs. This issue has been discussed in detail by a number of authors (e.g., Moll, et al., 1977; Fowler, et al., in press) and its general importance is reflected by the fact that it is taken up not only at this motor control symposium but in other papers (see MacNeilage's Status Report on Speech Production) and symposia at this IXth International Congress of Phonetic Sciences. Very briefly, the issue may be summarized as follows. Many investigators today contend that concepts which are relevant to the motor control of coordinated movements in general, whether from the walking movements of the hind leg of a cat or from arm movements about the elbow of a human being, are relevant also to the understanding of the motor control of the articulators for speech. It is argued that concepts related to the fine motor control of non-speech behaviors can and should be incorporated into speech production/motor control theory. In fact, I suspect that most investigators would accept such an argument, at least up to some specific point. That is, while many would agree that much fine motor control data from non-speech and from non-human research is of importance to speech production theory, they would also argue that in the end speech and language are distinctly human behaviors (although see MacNeilage's Status Report at this congress) and that the motor control of those behaviors is therefore unique, at least in some respects. For example, Bladon in his paper in this symposium, takes the view that "the physical facts of phonetics are at their most interesting when they serve to explain some aspect of phonology, to answer the question of why the sound systems of languages are the way they are." It is at this point that the impatience of many phoneticians becomes most evident, when they

note that in virtually all physiological/mechanical experiments on motor control mechanisms, correlates of abstract linguistic segmental units are conspicuous via their absence. Such units have proven extremely difficult to quantify. Thus, the question arises: should production theorists develop their own units and concepts which are based on actual experimental observations of motor control mechanisms in general and which are unbiased by notions and abstract concepts borrowed from linguistic theory? In the consideration of this question, either explicitly or implicitly, related questions and issues quickly arise. For example, Turvey and his associates (see, e.g. Fowler, et al., in press, for a review) describe much of modern phonetics research in production theory as consisting of "translation theories" designed to discover or elucidate the rules which could serve to translate from abstract linguistic units to the more concrete neurophysiological/mechanical data of speech motor control research. Turvey's use of Action Theory (Bernstein, 1967; Turvey, et al., 1978) and his development of the concept of "Coordinative Structures" represents an attempt to avoid such translation theories while at the same time not reject out of hand the use of all traditional linguistic concepts. The paper by Gay and Turvey in the present symposium provides some experimental consideration and discussion of the coordinative structures concept in speech motor control.

By its very nature, research in speech motor control, as exemplified by the reports in this symposium, is integral to issues such as these. Sussman, for example, discusses single motor unit behaviors and the insights they provide to temporal reorganization in coarticulation and to such prosodic events as stress, thus suggesting a means to provide "sensitive indicants of higher level linguistic conditions". Hirose provides data relevant to relationships between electromyographic activity and subsequent articulator movement. As MacNeilage points out in his status report paper at this congress, issues such as these cause questions concerning the role of feedback or closed loop control to become crucial. Indeed, the majority of the papers in this symposium at least refer to problems of feedback mechanisms while several specifically address themselves to such problems. Bladon proposes a "coarticulation resistance compiler" which is "linked

ambidirectionally" to satellite units in the motor control system. Abbs suggests a preliminary "multi-level control model" to account for observations of speech motor equivalence and compensatory articulation behaviors. Folkins also addresses the problem of motor equivalence, "functional interchangeability of activity level in different muscles", and compensations for mechanical modifications of articulator positioning. Perkell provides a discussion of recent compensatory articulation, or "bite-block", experimentation and thus the role of various sorts of feedback in speech motor control. He presents an example of the use of data from non-speech behaviors and in addition concludes that ideas such as those raised in motor control research are "closely related to questions about the nature of fundamental units which underlie the programming of speech production."

Thus, these papers on the Motor Control of Speech Gestures can be seen to confront some basic and crucial issues in phonetic theory. Further discussion of these and related issues is certain to bring us closer to an understanding of how it is that speech is generated and controlled.

References

- Bernstein, N. (1967): The coordination and regulation of movements, London: Pergamon Press.
- Fowler, C. A., P. Rubin, R. E. Remez, and M. T. Turvey (1978): "Implications for speech production of a general theory of action". In Language production, B. Butterworth (ed.), New York: Academic Press. (In press).
- Moll, K. L., G. N. Zimmerman and A. Smith (1977): "The study of speech production as a human neuromotor system" in Dynamic aspects of speech production, M. Sawashima and F. S. Cooper (eds.), 404-408, Tokyo: University of Tokyo Press.
- Turvey, M. T., R. E. Shaw, and W. Mace (1978): "Issues in the theory of action". In Attention and performance VII, J. Requin (ed.), Hillsdale, N. J.: Erlbaum (In press).

SPEECH MOTOR EQUIVALENCE: THE NEED FOR A MULTI-LEVEL CONTROL MODEL

James H. Abbs, Speech Motor Control Labs., University of Wisconsin, 1500 Highland Ave., Madison, Wisconsin 53706, USA

In the last ten years it has become increasingly difficult to view the neuromotor execution of speech as a series of descending motor commands, reflecting, in some direct manner, an underlying matrix of phonetic features. Rather, it appears that patterns of speech muscle activity may depend upon moment-to-moment peripheral conditions and adaptive modification of descending commands at several nervous system levels. In the present paper I would like to outline some current thoughts with regard to these speech motor processes, including some data from our laboratory and a preliminary model to account for recent observations.

If one considers speech motor control teleologically, the adaptive modification and adjustment of descending speech motor commands, based upon peripheral feedback, is quite appealing. For example, the orofacial system obviously serves the multiple functions of chewing, swallowing, and breathing, in addition to speech. Other less natural intrusions include cigarettes between the lips, chewing tobacco, a pipe between the teeth, etc. Because many of these activities can be performed simultaneously, without major interference or conscious compensation, a nervous system capability for on-line adaptive adjustment appears almost necessary. Such semi-automatic adaptation also seems likely in laryngeal and respiratory control as well. Recent physiological investigations of the laryngeal and respiratory systems, as well as consideration of their anatomy, indicate the profound influence that torso, head, and arm movements have upon the specific muscle contractions required for speech. Trained singers are quite aware of these influences. However, for many speaking situations, we have little difficulty sustaining continuous and intelligible speech concurrently with vigorous body movements. Observing a physical fitness teacher perform calisthenics and at the same time continuously cohort his or her pupils is an obvious example of this phenomenon. A preferable observation might be a cheerleader at a U.S. football game. In these and other similar cases, e.g., a vigorous university lecturer (an example suggested by Peter MacNeilage), one is impressed with our ability to produce continuous speech without major interference. Possibly these multiple concurrent motor

programs could be generated and pre-adjusted in parallel, but such an organization seems contradictory to the obvious availability of multiple afferent monitoring channels, documented differences in their nervous system origins, the principle of economy, and current information on normal and abnormal speech motor control.

In part these observations can be explained by the provocative model offered by MacNeilage (1970). He suggested that speech motor commands are adjusted to assure that individual articulators reach semi-invariant target positions, despite a substantial degree of variability in their starting positions. This kind of compensatory capability was referred to by Hebb (1949) as motor equivalence, although Hebb's definition was not quite so restrictive. Since MacNeilage's original paper, experimental observations have extended our appreciation of the adjustment capabilities operating in the speech motor control system. These recent observations appear to require an expansion of MacNeilage's insightful model and support the operation of motor equivalence in its most encompassing terms.

Indirect Experimental Evidence

The hypothesized operation of motor equivalence adjustments to descending speech motor commands implies a repetition-to-repetition flexibility in the way that a particular speech utterance is generated. Recent investigations support the operation of such flexibility both with regard to trade-offs between individual articulators and between individual synergistic muscles acting to move the same articulator. For example, it has been shown that the upper lip, lower lip, and jaw trade off reciprocally in their cooperative contributions to oral opening, viz., when the jaw had relatively large displacements the upper and lower lips had relatively small displacements, and conversely (Abbs and Netsell, 1973; Hughes and Abbs, 1976; Watkin and Fromm, 1978). Other investigators (Hasegawa et al., 1976) have reported lip and jaw reciprocity not only in regard to displacement, but for lip and jaw velocities as well. The trade-offs reported in these studies were observed for multiple repetitions of the same utterance where the net contributions of the individual movements (i.e., total oral opening or net velocity of closing) was relatively consistent. Comparable analyses of speech lung volume control illustrate a similar pattern of reciprocal trade-off between movements of the abdomen and thorax in producing subglottal air pressures (Hixon et al., 1973). In our

laboratory we have found other patterns of articulatory trade-off, including reciprocal interactions between the tongue and jaw (Chuang and Abbs, In Progress).¹

Not only do individual articulators appear to vary in their repetition-to-repetition contributions to a particular vocal tract objective, individual muscles appear to vary reciprocally in their combined contributions to an individual articulatory movement as well. In a recent experiment (Abbs and Kennedy, In Preparation), we found a reciprocal trade-off between the mentalis (MTL) and orbicularis oris inferior (OOI) muscles during repeated speech-related movements of the lower lip. This is in repetitions where the magnitude of MTL-EMG was relatively small, the magnitude of OOI-EMG was relatively large, and conversely. The flexibility of these adaptive speech motor command adjustments can be illustrated by considering this finding in relation to an earlier report by Sussman et al. (1973) of a parallel reciprocal trade-off between MTL-EMG magnitude and jaw lowering.

Overall these observations suggest that there may be several levels of programming and adjustment in the motor generation of speech. At some level, possibly corresponding to the phonetic feature input to the speech control system, overall vocal tract goals must be specified. However, due to the contrast between (1) the relative consistency with which these overall vocal tract goals are achieved, and (2) the variability of individual articulatory movements and muscle activity patterns, it would appear that these different output parameters are not programmed at the same levels of the nervous system.

Some Direct Evidence

The major issues with regard to this hypothesized motor control process concern the levels of the nervous system at which the adjustments might occur and the extent to which afferent feedback plays an important role. In attempts to more directly address these issues, several investigators have introduced unanticipated disturbances to the lips and jaw during ongoing speech (Bauer, 1974;

 (1) These patterns are most apparent in phonetically naive speakers. "Trained phoneticians" appear to produce speech, especially with regard to these reciprocal articulatory movements, quite different from that of normal speakers (cf. Gay, 1976).

Folkins and Abbs, 1975; 1976; Kennedy, 1977; Murphy and Abbs, In Progress). In these studies it was reasoned that if the nervous system sites where adaptive adjustments occurred were at "lower levels", semi-automatic, short-latency compensations would prevent unanticipated disturbances from interfering with ongoing speech. In the 15 subjects run with this particular paradigm, there have been no cases of disruption to ongoing articulation. In those studies where the latency of the compensations was discernible, it ranged from 25-50 msec. Compensatory responses have been observed in the muscles of the articulator to which the load was applied as well as in other articulators contributing to the same vocal tract goal, i.e., loads applied to the jaw yielded compensations in both the upper lip and lower lip musculature. The diffuse yet functional nature of these multiple compensatory responses corroborates the earlier suggestion that individual articulators and individual muscles can be adjusted flexibly to achieve desired overall vocal tract objectives. Based upon these findings, it appears that lower levels of the nervous system may be plausible sites for the adaptive modification of descending motor commands. Lower level sites for these adjustments are supported also by the observation that while the subjects in these studies perceived the articulator loading, they were unaware of generating the compensatory adjustments.²

A recent finding that might point to the possible origin of these compensatory adjustments is the observation that individuals with cerebellar disease and ataxic dysarthria are unable, without practice and conscious intervention, to adjust their lip movements to overcome experimental stabilization of the jaw (Netsell et al., In Preparation). Indeed, these patients report that many of their speech movements must be "consciously controlled". Certainly, if one accepts Eccles' (1973) suggestion, the cerebellum, with its multiple afferent and efferent connections, would be a primary candidate for yielding the semi-automatic, unconscious adjustments apparently required for normal speech. Other yet lower level sites

 (2) In our experience, the major difficulty in these unanticipated disturbance studies is discerning the peripheral manifestations of the disturbance. That is, while ongoing speech is seldom disrupted, the compensatory degrees of freedom are so great that one cannot always ascertain which muscles or movements were involved. This problem has apparently impeded some investigations using aerodynamic disturbances (Perkell, 1976).

might include areas in the brain stem where single point electrical stimulation yields very complex and semi-coordinated gestures of the laryngeal, masticatory, lingual, and facial musculature (Luschei, Personal Communication).

A Preliminary Model

Figure 1 is a schematic attempt to represent the motor control processes warranted from the data cited above.

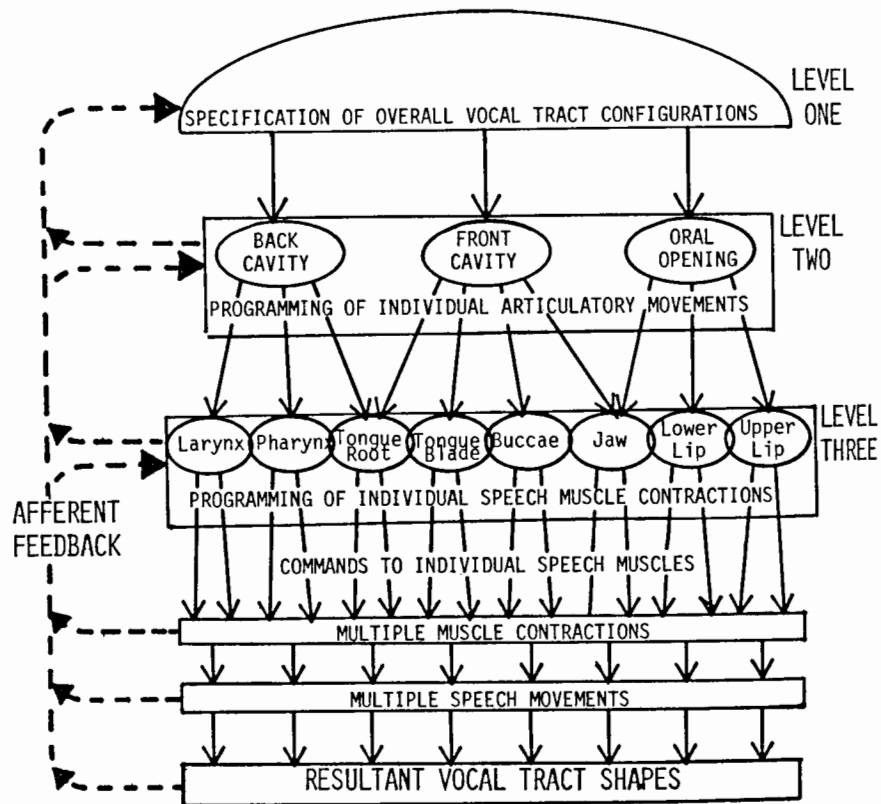


Figure 1

A multi-level model of the speech motor programming process. Solid lines represent descending control signals and dashed lines afferent feedback information. Dashed lines between levels of programming represent the ascending components of internal feedback pathways.

This model posits three levels of speech motor programming. At the highest level, overall vocal tract goals are specified, perhaps corresponding to some matrix of phonetic features. At the least, these goals represent the temporal-spatial configurations necessary for appropriate modulation of aerodynamic and acoustic signals. The second level of programming is involved in determining the particular set of individual movements that are to be employed in achieving the desired vocal tract goals. The third and final level of programming is concerned with specifying the individual muscle contraction patterns necessary to the generation of individual articulatory movements. These two lower levels of programming are based upon the observations (cited earlier) that (1) individual articulatory movements are not invariant with regard to particular vocal tract goals, i.e., repetitions of the same speech element, even if acoustically and perceptually similar, are produced often by different combinations of articulatory movements, and (2) individual muscle contractions are not invariant with regard to particular articulatory movements, i.e., repetitions of an articulatory movement are produced by different combinations of individual muscle contractions. As shown in Figure 1, it is posited also that the programming/adjustment of descending motor commands is accomplished with the aid of afferent feedback. This feature of the model is based upon the observations of compensatory responses to unanticipated articulator loading. That is, while it is plausible to consider parallel pre-adjustment of multiple motor commands (through some sort of efferent copy), in response to steady-state, anticipated disturbances (Lindblom et al., In Press), rapid adjustments to dynamic, unanticipated loads appear to require an afferent feedback control capability.

It is apparent from this representation that the model previously offered by MacNeilage does not account for all the motor command adjustments that apparently are accomplished by the speech motor execution system. That is, the adjustments to descending motor commands, at least as evidenced by the data cited above, obviously involve more than compensations for variations in individual articulator starting positions. Indeed, it appears that the primary controlled output parameters of the speech production system are not individual articulatory movements, but a series of overall vocal tract configurations. This model has other implications as

well. For example, analyses of individual articulatory movements or muscle contractions in relation to underlying phonetic features appear to be based upon the assumption that there is but a single level of speech motor programming. However, with multiple levels of adjustment, there is some question as to whether individual muscle contractions or articulatory movements are related, except in a probabilistic manner, to overall vocal tract phonetic features. If such a direct relationship exists, it may be necessary to hypothesize different features or to reallocate the current features, at least in part, to lower levels of the nervous system.

References

- Abbs, J. and R. Netsell (1973): "Coordination of the jaw and lower lip during speech production", ASHA Convention, Detroit.
- Bauer, L. (1974): "Peripheral control and mechanical properties of the lips during speech", M.S. Thesis, Univ. of Wisc., Madison.
- Eccles, J. (1973): The Understanding of the Brain, New York: McGraw.
- Folkins, J. and J. Abbs (1975): "Lip and jaw motor control during speech", JSHR 19, 207-220.
- Folkins, J. and J. Abbs (1976): "Additional observations on responses to resistive loading of the jaw", JSHR 19, 820-821.
- Gay, T. (1977): "Cine and EMG studies of articulatory organization", in Dynamic Aspects of Speech Production, M. Sawashima and F. Cooper (eds.), 85-102, Tokyo: Univ. of Tokyo Press.
- Hasegawa, A., M. McCutcheon, M. Wolf and S. Fletcher (1976): "Lip and jaw coordination during the production of /f,v/ in English", JASA, S84, 59.
- Hebb, D. (1949): The Organization of Behavior, New York: Wiley.
- Hixon, T., M. Goldman and J. Mead (1973): "Kinematics of the chest wall during speech production", JSHR 16, 78-115.
- Hughes, O. and J. Abbs (1976): "Labial-mandibular coordination in the production of speech", Phonetica 33, 199-221.
- Kennedy, J. (1977): "Compensatory responses of the labial musculature to unanticipated disruption of articulation", Ph.D. Thesis, Univ. of Washington, Seattle.
- Lindblom, B., J. Lubker and T. Gay (In Press): "Formant frequencies of some fixed mandible vowels and a model of speech motor programming by predictive simulation", JPh.
- MacNeilage, P. (1970): "The motor control of serial ordering in speech", Psych.Rev. 77, 182-196.
- Perkell, J. (1976): "Response to an unexpected suddenly induced change in the state of the vocal tract", MIT Res.Lab.Elect. 117, 273-281.
- Sussman, H., P. MacNeilage and R. Hanson (1973): "Labial and mandibular dynamics during the production of bilabial consonants", JSHR 16.
- Watkin, K. and D. Fromm (1978): "The control of labial movements by children", ASHA Convention, San Francisco.

MOTOR CONTROL OF COARTICULATION: LINGUISTIC CONSIDERATIONS

R.A.W. Bladon, Department of Linguistics, University College of North Wales, Bangor, U.K.

Various orientations to the motor control of speech

An orientation advocated recently by Moll, Zimmermann and Smith (1977) calls for priority to be given, in research on speech motor control, to exclusively neurophysiologically-based studies. The need is, they argue, to determine the properties of the human neuromotor system based on investigations of movement, muscle contraction and motor unit activity, freed from any constraints or a priori constructs imposed by linguistic considerations. Any processes or units of neuromotor coding which such enquiry were to establish might or might not subsequently turn out to correlate with linguistic units such as the phone, the feature or the syllable.

Without wishing to deny that the approach of Moll et al. has value, we propose to offer to this symposium the opposite orientation, wherein aspects of the descriptive linguistic apparatus are of prime importance. This decision reflects partly the conviction of a linguistic phonetician that the physical facts of phonetics are at their most interesting when they serve to explain some aspects of phonology, to answer the question why the sound systems of human languages are the way they are. The decision is also derived from the evidence that a wide range of phenomena of coarticulation are not obviously explainable (as yet, at least) in terms of the neuromotor system such as motor unit activity or articulatory velocity and inertia, but are referable to linguistically-defined entities which they thereby can validate.

In addition, many models of the speech production processes seem to occupy the middle ground between these two extreme positions. Among the proposals which might be grouped together here are those of Kozhevnikov and Chistovich (1965), Henke (1966), MacNeilage (1970), Gay (1977a) and Perkell (1977). In very general terms, these models are of a basic "wedding-cake" form such as Figure 1: that is, they are arranged sequentially as tiered boxes, with a distinct top and bottom corresponding to mechanisms associated respectively with more central cortical functions and with more peripheral ones. The number of tiers, and the content of each one, is stylised and is not meant to be attributed specifically

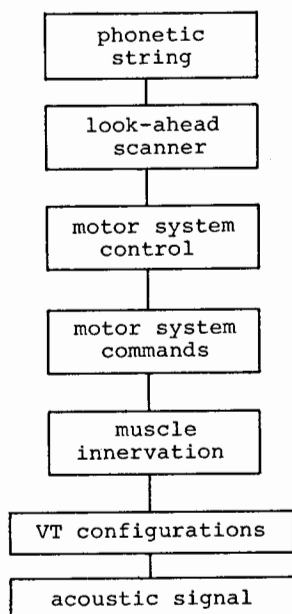


Figure 1

Typical "wedding-cake" structure of some speech production models

control system for coarticulation is organised not in a unidirectional, tiered fashion, but as a set of satellites (many representing linguistic factors) linked ambidirectionally to a nuclear body, the CR (Coarticulation Resistance) compiler. The following formulation is a brief summary of this theoretical position. Justification of it, of an organisation which reduces the emphasis on the supposed sequential nature of the speech processors and which minimises the "more central/less central" distinction, and of the particular components recognised in the model, has been given elsewhere (Bladon and Al-Bamerni, 1976; Bladon, 1978).

Information relevant to the direction and domain of coarticulatory effects seems to derive from a wide variety of satellite sources, some quasi-universal (such as the boundary of an intonation-group, which is widely observed to impede the temporal spread of coarticulated features), some language-specific (such as the

to any author. Varying types of feedback between tiers are postulated, but not shown in the figure. A possible objection to the sequential arrangement is that, while no doubt well motivated for the lowest tiers representing the transduction of speech between various fairly accessible transmission channels, such an arrangement is more speculative as a claim about the higher levels of the central nervous system.

The "wedding-cake" models share with our own orientation (and in contrast to Moll et al.) an interest in the linguistic nature of the input, which they normally state to be a string of discrete phonological units.

By contrast with the claims implied by Figure 1, however, our conception of the "upstream" processes feeding data to the motor

report by Ladefoged (1967) that while French and English both show a /k/ coarticulatorily advanced before an /i/ vowel, only French shows the similar effect after /i/), and some speaker-sensitive. Our discussions of these factors have led Kent and Minifie (1977, 120) to write: "Perhaps the solution to coarticulation is as complex as this multiplicity of factors suggests, but ... (they) represent the contributions of many unknown, or poorly known effects". This comment is valid; but it is not a criticism of our position. It seems to be inescapable that the control of coarticulation in speech is indeed governed by a multiplicity of factors.

With a view to integrating these disparate factors in a theory, let us initially postulate the notion of coarticulation resistance (CR) as the central principle of articulatory control. The speech production mechanism is hypothesised to have continuous access to CR information, which can be considered to attach to each allophone and phonetic boundary. It is important to realise why an initial CR specification is tabulated for each allophone. A classic demonstration of this is afforded by the RP English /l/ allophones, of which dark syllabic [ɫ] ('fiddle') is highly resistant to coarticulation, dark nonsyllabic [ɫ̥] ('feel') is somewhat less so, and clear nonsyllabic [l] ('leaf') is very much less resistant. We are aware of no explanation of this behaviour in any terms, linguistic or neurophysiological: the idiosyncrasies can only be handled by assuming something like an allophone-specific assignment of a CR value. This numerical value is re-computed at a level of articulatory planning by the CR compiler to take account of the wide range of relevant coarticulatory constraints.

In what follows we presuppose a phonological apparatus broadly of the kind of Chomsky and Halle (1968). Within their phonological component, various kinds of linguistic construct will be examined, in conjunction with the control mechanisms related to them.

Phonological constructs related to units of articulatory planning

Many early coarticulation studies hoped to identify a single determinant regulating the control of the domain of coarticulatory effects: an invariant unit of production which would have a linguistic counterpart, such as the phone, syllable or phonetic feature. The hope was a vain one. The weight of evidence now available suggests that the coarticulatory control mechanism is sensitive not to any one invariant unit alone but, at different times, to (at least) all those three.

The phonetic feature is at the basis of coarticulation theory in that typical cases of coarticulation arise by definition from the asynchrony of events associated with different articulators. This is reflected phonetically in the temporal spreading of a feature. It is true that the speech control mechanism can be highly sensitive to the feature being coarticulated. Thus, for example, it has been shown that English /s/ occurring in CCC clusters blocks the spread of anticipatory jaw-opening before /æ/ (Amerman, Daniloﬀ and Moll, 1970); and that /s/ resists any shift in its tongue-bladeness (towards a tip articulation) adjacent to /t d n l/ (Bladon and Nolan, 1977); but that this resistance to coarticulate is specific to the coarticulated feature in question, because /s/ freely allows coarticulated labialisation anticipating an /u/ vowel (Daniloﬀ and Moll, 1968). It is equally true that to propose the feature as the sole unit of coarticulatory control would be unattractive, as it would not account for example for British English clear [ɪ], which is quite free in its coarticulation in respect of any of the features vowel-quality, lateral-quality and voicelessness indiscriminately (Bladon and Al-Bamerni, 1976).

Numerous cases, such as the last-mentioned, argue for the phoneme (or perhaps better, the extrinsic allophone) as the unit of articulatory planning. Two further examples may be mentioned. In Italian, intervocalic consonants demonstrate an equal degree of coarticulated tongue-body movement with both a preceding and a following vowel, thus irrespective of syllable boundaries. In French, the anticipatory spread of velum lowering before a nasal, as revealed by EMG, is over a limited domain within a string of preceding oral vowels (Bladon and Carbonaro, 1978; Benguerel et al., 1977). Such arguments for the allophone-sized unit tend, however, to be of a "default" kind, postulated whenever coarticulation fails to coincide with syllable boundaries in some sense. Generalising the allophone, in the interests of proposing an invariant unit, to cases which the syllable could have successfully delimited, has led to an overall too weak hypothesis concerning coarticulatory domain, such as that of Henke's model (1966), which predicted coarticulatory activity whenever a segment showed no antagonistic specification. Our model avoids this problem by two expedients: first, by a segment-specific index of CR (referred to earlier) which inhibits coarticulatory spread in appropriate circumstances, and second, by recognising a plurality of articulatorily-relevant

units which will include the syllable as required.

The phonological syllable, neglected by Chomsky and Halle, has since 1968 enjoyed a revival. Syllable-structure rules in phonology would define the syllable differently for different languages; nevertheless, the structure CV has a claim to universal preference in that, first, there appear to be no languages without CV syllables, second, several languages have syllables of only the CV type, and third, CV is the attested structure in early language acquisition. We profoundly disagree, therefore, with Gay's opinion (1977a) that Kozhevnikov and Chistovich's notion of an articulatory syllable of the form C_0V (where C_0 stands for any number of consonants) is "an unnatural and counterintuitive syllable that bears no simple correspondence to common linguistic or phonetic units." Within their articulatory syllable, it will be recalled, coarticulation was hypothesised to be maximal. A great deal of evidence supports this hypothesis, notably the labialization of a string of C before /u/ in Russian (Kozhevnikov and Chistovich, 1965), in English (Daniloﬀ and Moll, 1968) and in French (Benguerel and Cowan, 1974); and also the finding (Bladon, 1977) that even the relatively weaker lip-rounding accompanying English /r/ extended leftwards to the same C_0V boundary.

Other substantive constructs in phonology

Explanation of the control of coarticulatory behaviour in VC positions has remained elusive. Relevant data here include the anticipatory nasalisation of English vowels before nasals (Moll and Daniloﬀ, 1971); American English /r/ which coarticulates with adjacent vowel quality more readily in the final position than in the initial CV position (Lehiste, 1964); or, in VCC sequences, the consonantal influence upon tongue apex position in V (Amerman and Daniloﬀ, 1977). Current phonological theory suggests an explanation in terms of the phonological strength hierarchy. Based on a variety of evidence including sound-change, phonological segments, sequences and positions in the word are assigned a degree of phonological strength. VC positions are weak, since they show more phonological assimilations and elisions. It is reasonable to suggest that the coarticulatory control mechanism is sensitive to this, as to other linguistic properties.

A second such property is the lexical representation of the inventory of phonological items in a language. The degree to which a lateral, for instance, undergoes vowel-quality coarticulation

varies according to the number of laterals in a language's phonological system: in our data, Irish, with three laterals to be kept distinct, shows very little quality coarticulation in comparison with American English, with only one (but highly coarticulated) lateral; Swedish or Italian, with two laterals each, fall in between with respect to coarticulation. The need in such cases to maintain phonemic distinctions (short of the point of incipient sound change and phonemic restructuring) has widely been held to place an upper bound on the extent of coarticulatory behaviour. The principle appeared to make the wrong predictions in the data of Benguerel and Cowan (1974), however, who found that lip protrusion anticipating French /u/ could sometimes extend transconsonantly into the preceding vowel, despite the apparent threat to the lexical contrast /i - y/ in French.

It seems certain that rapid-speech variations are subject to a degree of coarticulatory control. Gay (1977b) showed that at a fast speaking rate a vowel F2 transition effectively begins at a point of greater overlap with the preceding consonant than at a normal speaking rate. Rapid or casual speech variants are coming under scrutiny by phonologists in order to validate their substantive hypotheses of rule ordering. In the derivation of the (ultimate?) rapid-speech form [də.viː] 'divinity', Stampe (1972) demonstrates fairly convincingly that phonological processes do not apply in a linear order, but whenever the configurations they would eliminate arise. Among the processes concerned is the coarticulatory one of intra-syllable vowel nasalization, which re-applies three times, as successively more rapid forms are derived. This cyclic manner of application has important implications for the operational design of the motor control component of the speech production processes, and strongly supports the notion of ambidirectional, on-line exchange of information between the CR compiler (or its equivalent) and the linguistic rule system.

The testing of these various elements of the coarticulatory model by predicting from them onto new data, turns out to be partly successful, but, as has been demonstrated for several cases, partly unsuccessful. Apparently, no one linguistically-related mechanism will explain all or even a majority of observed coarticulatory behaviour. How to assign a weighting to the separate contribution of each mechanism, and indeed how many such mechanisms there are, remain research questions for the future.

References

- Amerman, J.D. and R.G. Daniloff (1977): "Aspects of lingual coarticulation", *JPh* 5, 107-114.
- Amerman, J.D., R.G. Daniloff and K. Moll (1970): "Lip and jaw coarticulation for the phoneme /æ/", *JSHR* 13, 147-161.
- Benguerel, A.-P. and H.A. Cowan (1974): "Coarticulation of upper lip protrusion in French", *Phonetica* 30, 41-55.
- Benguerel, A.-P., H. Hirose, M. Sawashima and T. Ushijima (1977): "Velar coarticulation in French", *JPh* 5, 159-168.
- Bladon, R.A.W. (1978): "Some control components of a speech production model", in *Current Issues in the Phonetic Sciences*.
- Bladon, R.A.W. and A. Al-Bamerni (1976): "Coarticulation resistance in English /l/", *JPh* 4, 137-150.
- Bladon, R.A.W. and E. Carbonaro (1978): "Lateral consonants in Italian", *Italian Linguistics*, forthcoming.
- Bladon, R.A.W. and F.J. Nolan (1977): "A videofluorographic investigation of tip and blade alveolars in English", *JPh* 5, 185-193.
- Chomsky, N. and M. Halle (1968): *The Sound Pattern of English*, New York: Harper & Row.
- Daniloff, R.G. and K. Moll (1968): "Coarticulation of lip-rounding", *JSHR* 11, 707-721.
- Gay, T. (1977a): "Articulatory units: segments or syllables?", *Paper read at Symposium*, Boulder, Colorado.
- Gay, T. (1977b): "Articulatory movements in VCV sequences", *JASA* 62, 183-193.
- Henke, W. (1966): *Dynamic articulatory model of speech production using computer simulation*, Ph.D. thesis, M.I.T.
- Kent, R.D. and F.D. Minifie (1977): "Coarticulation in recent speech production models", *JPh* 5, 115-133.
- Kozhevnikov, V.A. and L.A. Chistovich (1965): *Speech: Articulation and Perception*, *JPRS* 30, 543, US Department of Commerce.
- Ladefoged, P. (1967): *Linguistic Phonetics*, UCLA Working Papers in Phonetics 6, Los Angeles: UCLA.
- Lehiste, I. (1964): *Acoustic characteristics of selected English consonants*, Bloomington: Indiana UP.
- MacNeilage, P.F. (1970): "Motor control of serial ordering of speech", *Psych.Rev.* 77, 182-196.
- Moll, K.L. and R.G. Daniloff (1971): "Investigation of the timing of velar movements in speech", *JASA* 50, 678-684.
- Moll, K.L., G.N. Zimmermann and A. Smith (1977): "The study of speech production as a human neuromotor system", in *Dynamic Aspects of Speech Production*, M. Sawashima and F.S. Cooper (eds.), Tokyo.
- Perkell, J. (1977): "Articulatory modeling...", in Carre R. et al. *Modèles articulatoires et phonétiques*, G.A.L.F., 197-192.
- Stampe, D. (1972): *A dissertation on natural phonology*, Ph.D. thesis, U. Chicago.

SEGMENTAL INVARIANCE RECONSIDERED

R.G. Daniloff, Purdue University, W. Lafayette, Indiana, USA
 M.A.A. Tatham, Languages and Linguistics, University of Essex,
 Colchester, UK

Stop consonant articulation is context sensitive, Fromkin (1966). Such context sensitivity, often related to coarticulation, is taken as support for complex, high level "encoding", Kozhevnikov and Chistovich (1965). The duration and force of labial closure have been variously shown to be context sensitive to (1) syllable position, (2) voicing, (3) stress and (4) vocalic context. We concluded that previous cinefluorographic, electromyographic, and palatographic studies may have overestimated the extent of context sensitivity by failure to control for tempo, speech effort, task learning, and sentential context. The purpose of the study reported herein was to reassess the context sensitivity of the EMG impulse associated with labial-stop closure.

Procedures

4 adults served as subjects. Each was required to repeat previously tape recorded sentences as he heard them over earphones. The tape recorded sentences were carefully controlled for constant tempo, stress levels, and good phonetic quality. Speech tokens consisted of all combinations of C_xVC , CVC_x , C_xVC_x , 'VCV, and $V'CV$ tokens, $C_x = [p,b]$, $C = [d]$, $V = [i,u,\lambda]$. CVC items were spoken in the sentence, "He'll spoof the [CVC] again", and VCV items in, "Smell this poof of [VCV] again", such that each token received primary, sentence level stress. Subjects visually monitored their vocal output as they spoke, keeping it at 60-65 dB, SPL. Each subject practiced the task of repeating the tape recorded sentences with as controlled a tempo, vocal output, and desired stress level as possible.

Bipolar silver disc surface electrodes on the upper lip served to detect the EMG pulse on the orbicularis oris for the lip closing gesture for [p,b]. The EMG and Raw Voice signals were recorded, rectified, and integrated, and on a single, 5 channel mingograph output, raw/integrated EMG, voice, and timing signals were displayed. The upper lip-surface electrode array was chosen because: (1) the upper lip is accessible, (2) labial surface EMG signals are quite interpretable-reproducible, (3) myo- and biomechanically

the lip in its closing gesture is a simple system, (4) the labial closure gesture has been extensively studied, (5) surface electrodes potentially yield a better estimate of whole-muscle activity than do needle/hooked wire electrodes, (6) labial closure is reputedly context sensitive.

Subjects received extensive practice beforehand; subjects who could not relax the lips to nearly zero-baseline EMG activity between utterances were rejected. Each of 40 tokens was repeated 26 times for a total of 1040 repeated, randomized sentences spoken at one sitting, with pauses every 10 minutes for relaxation.

Results:

The results of the study are based upon 4 criterion measures: the peak amplitude of the EMG pulse for lip closure, the duration of the EMG pulse, the delay between EMG onset and acoustic burst release, and the delay between peak EMG level and burst release. Data were analyzed using 3 way ANOVAs, with a conservative $\alpha \leq 0.1$. The concise results are shown in the table below. Interpretation of the results is based upon the following assumptions: lip closure for stops is mediated primarily by orbicularis oris contraction (00). M.O.O. force of contraction and the height of the integrated EMG signal are linearly, if not monotonically related. If one of the particular aspects of context-sensitivity being investigated were a part of speakers' linguistic competence, it was our expectation that all 4 speakers should show a statistically significant and similar shift in the labial closure criterion measure associated with that aspect of context since by our estimate, 25 repetitions of each token offered a firm basis for statistical inference.

The peak EMG amplitude measure presumably relates directly to maximum force of M.O.O. contraction. As shown, in every case, one or more subjects for one or more tokens showed a non-significant change in EMG peak amplitude; and in fact, for all 4 contextual effects: voicing, syllable position, vowel, stress - at least one subject showed a reversal of trend, with differences in EMG peak being just opposite those shown for 2 or 3 of the subjects. Thus, there is only a modest trend for voiceless stops in /i/ context to be modestly more effortful, muscularly. Stress and syllable position had no consistent effect upon peak EMG amplitude. Duration of the EMG pulse revealed a strong dependence upon context such

	VOICE	POSITION	VOWEL	STRESS
Peak Amplitude EMG	-voice>voiced; moderate effect; subject, token dependent	initial>final; weak effect; very subject, token dependent	/i/>/u/>/u/, moderate effect; subject, token dependent	not consistent; weak effect; strongly subject dependent
Duration of EMG Pulse	-voice>voiced; strong effect; little subject or token dependence	initial>final; strong effect; small subject, token dependence	not consistent; moderately weak effect; very token, subject dependent	V2 stress>V1 stress; strong effect; small subject dependence
Delay EMG Onset to Burst	voiceless final>voiced final; moderate effect; token, position dependent	initial>final; strong effect; small token, subject dependence	/u/>/i/ weak effect; strongly token, subject dependent	V2 stress>V1 stress; strong effect; small subject dependence
Peak EMG to Burst Onset	voiceless>voiced, moderate effect; subject, token dependent	voiced initial>voiced final; moderately strong effect; strong voicing, moderate subject dependence	not consistent, weak effect; strong token, subject dependence	V2 stress>V1 stress; moderate effect; small subject dependence

that voiceless stops were longer than voiced, initial stops were longer than syllable final stops, and pre-stress-position stops were longer than post-stress stops. In all three cases, tokens for one subject failed to achieve significance. In addition, for the voicing effect, one subject's voiced stops were significantly longer than his voiceless tokens. For the time delay between EMG onset and burst release, the effect of syllable position was fairly strong in that all initial stops began earlier and, in 6 of 8 cases, the difference was significant. For voicing, there was a modestly strong trend for voiceless stops to begin earlier, in 11 of 12 cases, with 10 of the 12 cases being significantly greater. Stress had a strong effect in that the pre-stress stop began earlier for all 4 subjects, significantly so in 3 of 4 cases. Vowels had no consistent effect upon delay. The peak EMG to acoustic release temporal delay measure shows a weak dependence upon voicing; the effect of the vowel upon delay was weak and inconsistent. The effect of stress upon this measure was only moderate in that for all subjects, pre-stress stops had earlier occurring peaks, but these differences were significant in only two of four subjects. The effect of position upon this delay measure was moderately strong in that in all 8 cases, syllable initial vowels had earlier occurring EMG peaks, and in 6 of 8 cases, the differences were significant.

Conclusions

Contrary to the work of Fromkin, syllable final stops were shorter in all cases, significantly so in 6/8 cases, than initial stops. Syllable position had no consistent effect upon the amount of muscle activity for closure, but initial stops began earlier, vis-à-vis onset or peak of EMG and burst release, in all cases, and significantly earlier in 12 of 16 cases. Thus, syllable initial stops are generally longer and earlier in onset, but not muscularly more effortful. Vowel context effects were enigmatic. It was expected that one would have earlier and stronger EMG pulses for [i] than for [ʌ] than for [u]. This was not the case; in the majority of cases, vowel context had non-significant effects on most EMG measures, and even when significant, the direction of the trend varied from subject to subject. Stress was potent as a factor such that pre-stressed stops began earlier in all cases, significantly so in 5 of 8 cases, and in 3 of 4 cases, the EMG

pulse was significantly longer. However, stress had no systematic effect upon the amount of muscle activity needed for closure. Finally, voice as a factor had no systematic, cross subject effect upon amount of muscle activity. With only one reversal (significant), in 9 of 12 cases, voiceless stops were longer than voiced stops, and EMG onset began earlier, vis-à-vis release, in 10 of 12 cases, one exception being a significant reversal. The effect of voice upon the EMG peak to burst release measure was highly variable.

The most startling result was that without exception, at least one subject, for at least one token showed either a non-significant context effect, or a reversal of trend for a given context effect. According to a strict criterion of all subjects and all tokens revealing a significant change in criterion measure, then, not one of the contextual effects investigated is less than idiosyncratic, i.e.: the contextual effects are a trend, but not absolutely a component of linguistic performance. Further analysis showed that subject sex and naïveté had no effect upon the results. Surprisingly, the phonetic shape of the syllable, e.g.: C_xVC_x vs. CVC_x vs. VCV had a profound effect upon all criterion measures, and was probably the single most potent effect found in this study. It is difficult to explain why vowel context did not produce more, and more systematic changes in the size and timing of the EMG patterns for labial closure. It may be the case that coarticulation of stops and vowels, known to be quite a strong effect, is highly idiosyncratic. Or, it may be the case that the transformation of muscle activity into final vocal tract shape is complex, and non-linear, within and across subjects, so that interpretation and comparison of EMG data are more complex than is suspected. It is our conclusion that labial closure as an articulatory gesture is relatively context insensitive as far as amount of muscle activity is concerned. It is context sensitive as far as syllable position, voicing and stress are concerned in that voiceless, initial, pre-stress stops are generally longer and begin earlier: however, certain subjects and tokens violate this trend. We conclude that electromyographic signals, especially vis-à-vis coarticulation, may be more complex to interpret than is presently suspected.

References

- Fromkin, V.A. (1966): "Neuromuscular specification of linguistic units", L&S 9, 170-199.
- Kozhevnikov, V. and L. Chistovich (1965): Speech: articulation and perception, JPRS 30, 543, U.S. Bureau Commerce, Washington, D.C.

MASSETER, TEMPORALIS, AND MEDIAL PTERYGOID ACTIVITY WITH THE
MANDIBLE FREE AND FIXED

John W. Folkins, The University of Iowa, Iowa City, Iowa, U.S.A.

Experiment One

Introduction

In general each of the speech articulators is acted on by a number of anatomically different muscles. The muscles which act on a given articulator may either 1) have activity levels which are functionally interchangeable or 2) each muscle may have a separate function which is seldom (if ever) accomplished by substitution of activity in other muscles. The Russian physiologist, Nicholas Bernstein (1967), has stressed the extent to which the first possibility; i.e., the functional interchangeability of activity level in different muscles, operates in normal human movements. In relation to speech, MacNeilage (1970) presents a perspective which (in this one respect) is related to Bernstein's ideas. MacNeilage believes there is a ubiquity of variability in muscle activity for attainment of vocal tract targets.

Textbooks of speech anatomy (Palmer, 1973; Zemlin, 1968) imply that masseter, temporalis, and medial pterygoid act in a similar manner to raise the mandible during speech. If this is the case, these muscles might typically operate interchangeably in many combinations for similar speech movements. However, the jaw closing muscles have been studied extensively in the dental literature (e.g., Kawamura, 1974) and the anthropological literature (e.g., Hylander, 1975). These studies have illustrated important differences in function and activity patterns between jaw-closing muscles. On the basis of these studies one might expect that each jaw-closing muscle has a specific functional role during speech and its activity is not typically interchanged with activity in other muscles.

There is not room to discuss the electromyographic (EMG) studies of the jaw-closing muscles during speech. However, the research to date does not provide much data concerning the above issue. Therefore, the purpose of the present study is to examine EMG activity and make comparisons between and within jaw-closing muscles during speech.

Method

Hooked-wire electrodes were used to record EMG from masseter, temporalis, and medial pterygoid in four normal adults. Jaw movement was transduced with a strain gauge technique. Each subject produced three to six repetitions of 11 isolated syllables, seven syllables in a carrier phrase, trains of syllables at various rates, and the rainbow passage. The limitations on the length of this paper preclude adequate description of experimental methods; however, a full report of this research will be ready for publication soon.

Results and Discussion

Figure 1 shows a typical example of EMG activity during the first sentence of the rainbow passage. Medial pterygoid was the most active muscle, not only in this example, but for all four subjects in almost all speech tasks. This example is typical as medial pterygoid tends to be: 1) moderately active throughout the utterance, 2) most active in relation to jaw closing, and 3) reduced during jaw opening. In Figure 1 masseter and temporalis were slightly active, but in many instances they were quiet throughout the speech sample. When masseter and temporalis were active, it tended to be during jaw-closing movements of large displacement or velocity.

Even though one might hope medial pterygoid activity would increase when the jaw moves further or faster, this is not necessarily the case due to the nonlinear relations between EMG and muscle force (Bigland and Lippold, 1954), and the difficulty in relating jaw-closing forces to parameters of jaw movement. As illustrated for isolated VCs by one subject in Figure 2, medial pterygoid EMG did not increase as a function of displacement. In fact, for the four subjects half the correlations between peak medial pterygoid EMG and displacement (0.15, -0.25, -0.37, and -0.17) and peak velocity (0.53, 0.23, -0.32, and 0.14) were negative. If one assumes that more muscle force is required to increase jaw-closing displacement or velocity, then either this is not reflected in our EMG measurements or is produced by muscles other than medial pterygoid. Figure 2 also shows that for isolated VCs temporalis became more active for larger displacements ($r = 0.66$). Three of the four subjects tended to increase

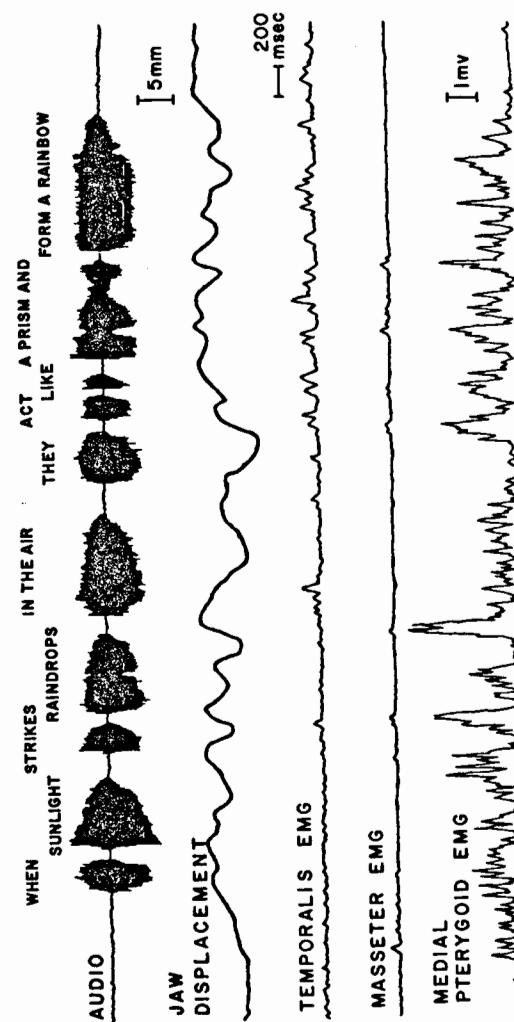


Figure 1. Rectified and smoothed (20 msec TC) EMG during reading for subject 3.

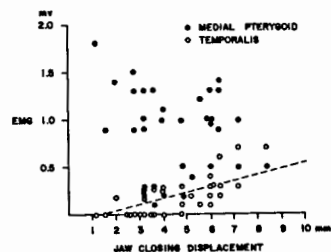


Figure 2. Scatterplot of peak EMG and jaw-closing displacement for isolated VCs by subject 4. The dashed line is a linear regression through the temporalis points. Masseter was zero for all syllables.

temporalis activity for larger (or faster) jaw-closing movements. The other subject tended to use masseter rather than temporalis.

A notable aspect of Figure 2 is the spread in EMG values for movements with similar displacements. Variability is especially evident in repeated syllables with matched displacement and velocity. For example, the subject in Figure 2 repeated [pæ] 24 times in rapid succession. Closing displacement was consistent as one standard deviation was only 16% of the average displacement. One standard deviation of peak velocity was only 14% of average. Mean medial pterygoid activity was 1.05 mv; however, it showed a standard deviation of 43%. Masseter averaged 2.15 mv with a standard deviation of 81%. This is surprising as masseter was quiet throughout the isolated syllables for this subject even though the repeated syllables were well within the range of displacements and velocities for isolated syllables. Variability was evident for most situations, but occasionally subjects were consistent. For example, on the left of Figure 3 one standard deviation of peak medial pterygoid EMG is only 10% of the mean.

In summary, medial pterygoid is consistently more active than masseter and temporalis. However, within this general distinction there appears to be a large amount of utterance-to-utterance variability in the way these muscles are employed.

Experiment Two

Introduction

A number of papers have illustrated the abilities of the speech motor control system to compensate for mechanical modifications in movement of the jaw (Folkins and Abbs, 1975; Lindblom, Lubker, and Gay, in press). Both Lindblom et al. and Perkell (1979) suggest that speech motor systems produce appropriate gestures in spite of perturbing factors by employing central stimulation strategies. For example, when one speaks with a bite block, a central movement plan adjusts the roles of many articulators for the lack of jaw movement. As the jaw is fixed with the bite block one might also expect the central movement plan to eliminate "unnecessary" jaw-closing muscle activity. The purpose of this experiment was to record EMG from the jaw-closing muscles with a bite block in place and see if there is a reorganization of muscle activity.

Method

This experiment was carried out in the same experimental sessions, with the same electrode placements as experiment one. After producing the speech sample with the jaw free to move, the sample was repeated with the jaw fixed with a bite block. Bite blocks providing both 5 mm and 15 mm of interincisor distance were employed.

Results and Discussion

With both sizes of bite blocks there were consistent bursts of EMG activity which related closely to the temporal patterns of EMG found in each muscle during the jaw free condition. This is illustrated in Figure 3 for medial pterygoid as [pæ] was repeated at a fast rate.

As the mandible is not moving, it is not clear why the phasic jaw muscle activity persists with the bite block. A complete central reorganization would be expected to remove unnecessary muscle activity. As an alternative, it may be that the phasic jaw muscle activity is involved in the organization of other articulatory movements occurring with the bite blocks. That is, peripheral motor control mechanisms (including brainstem reflexes; McClean, Folkins, and Larson, in press) may be important

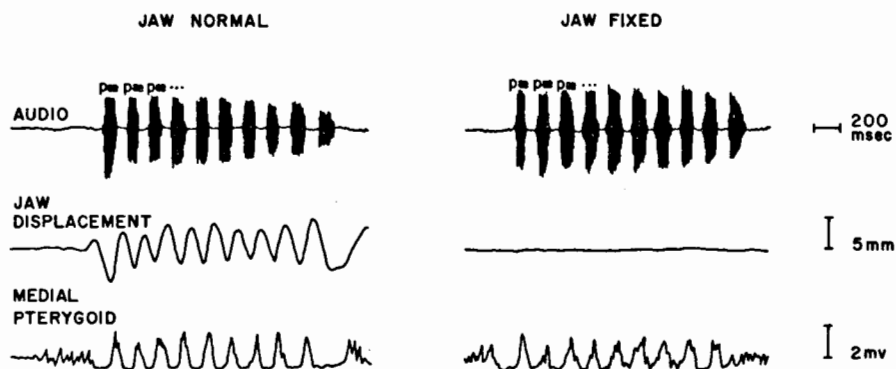


Figure 3. Rectified and smoothed (20 msec TC) EMG for [pæ] repeated at a fast rate by subject 1.

components in the processes which accomplish appropriate speech movements with and without mechanical interferences.

References

- Bernstein, N. (1967): The Coordination and Regulation of Movements, New York: Pergamon Press.
- Bigland, B. and O. Lippold (1954): The relation between force, velocity, and integrated electrical activity in human muscles, J. Physiol. 123, 214-224.
- Folkins, J. and J. Abbs (1975): Lip and jaw motor control during speech: Responses to resistive loading of the jaw, JSHR 18, 207-220.
- Hylander, W. (1975): The human mandible: Lever or link?, Am. J. Phys. Anthrop. 43, 227-242.
- Kawamura, Y. (1974): Physiology of Mastication, Basel: S. Karger.
- Lindblom, B., J. Lubker, and T. Gay (in press): Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation, J. Phonetics.
- MacNeilage, P. (1970): Motor control of serial ordering of speech, Psych. Rev. 77, 182-196.
- McClellan, M., J. Folkins, and C. Larson (in press): The role of the perioral reflex in lip motor control for speech, Brain and Language.

Palmer, J. (1972): Anatomy for Speech and Hearing, 2nd Ed., New York: Harper & Row.

Perkell, J. (1979): Phonetic features and the physiology of speech production, in Language Production, B. Butterworth (ed.), New York: Academic Press.

Zemlin, W. (1968): Speech and Hearing Science: Anatomy and Physiology, Englewood Cliffs, New Jersey: Prentice Hall.

EFFECTS OF EFFERENT AND AFFERENT INTERFERENCE ON SPEECH PRODUCTION:
IMPLICATIONS FOR A GENERATIVE THEORY OF SPEECH MOTOR CONTROL

Thomas Gay and Michael Turvey, Haskins Laboratories, New Haven, Connecticut, U.S.A.

One might claim that speech production proceeds in open-loop fashion: for a given speech sound a motor program prescribes a standard set of instructions to the musculature. Against this claim, however, is the fact that the backdrop of articulatory states into which the standard instructions would be inserted is itself not standardized. The initial conditions (or contexts) for the articulatory gestures yielding a given speech sound vary considerably (cf. MacNeilage, 1970). This is a most notable feature of speakers: within reasonable limits they are capable of producing the necessary configurations of articulatory maneuvers for the sounds of speech even though the departure points for those configurations are ever-varying. Moreover, it appears that the configuration of gestures underlying a desired speech sound can be generated with virtually no experimentation and without the benefit of auditory monitoring. Lindblom and his co-workers (Lindblom et al., 1978) have shown that speakers fitted with bite blocks can produce isolated vowels within the range of variability for normal vowel production, and that satisfactory formant matching occurs on the first pitch pulse of the first attempt.

Clearly, the adaptive, generative nature of articulation is not captured by the notion of open-loop control. Consequently, speech investigators have turned to the claim that the control of speech is closed-loop. In closed-loop explanations, a sensory referent is proposed that relates either to the environmental goal of the articulatory gestures, such as a spatial target or an acoustic pattern, or to the movement-producing commands. (The interpretation of sensory referent as a spatial target is currently the more popular interpretation.) The comparison of sensory feedback with the sensory referent yields an error signal that provides the basis for adjusting the lower level motor mechanism(s) responsible for controlling the referent. Over successive comparisons an increasingly closer match between the feedback sensory signal and the desired sensory referent is achieved.

While a closed-loop mechanism can, in principle, adjust motor instructions to variable initial conditions to attain the referent, it is not immediately obvious how a feedback mechanism that gradually approaches a desired result could underly the immediate adjustment to context evidenced in everyday speaking and underlined by the phenomenon reported by Lindblom and his colleagues. What is needed is a mechanism that: (1) can produce the appropriate articulatory gestures in the face of variable and often novel initial conditions, and (2) can do so without trial and error.

On first thought, these two criteria are met by model-referenced control. Here, the closed-loop mechanism tied to the peripheral speech apparatus is modeled centrally so that motor commands and their sensory consequences can be simulated for the current conditions of the peripheral speech apparatus. The simulated motor commands that result in a match between the simulated sensory feedback and the sensory referent are then realized as actual motor commands. In principle, the predictive simulation of model-referenced control could underly the immediate readjustment phenomenon (Lindblom et al., 1978). There is, however, a potentially serious drawback to any closed-loop explanation: While an error signal can index how near the collective action of a number of muscles is to the desired consequence, it does not prescribe in any straightforward way how the individual muscles are to be adjusted to give a closer approximation to the referent (Fowler and Turvey, in press).

There is another mechanism, very different from closed-loop control, that meets the two criteria noted above. The rationalization and evidence for this mechanism - referred to as a coordinative structure - has been presented elsewhere in some detail (Fowler, 1977; Fowler, Rubin, Remez and Turvey, in press; Turvey, Shaw and Mace, in press). A rough sketch must suffice for current purposes.

Consider a set of several (relatively) independent muscles. As an aggregate, the muscles would exhibit a large number of degrees of freedom and would rely on a source external to themselves for their control. The number of degrees of freedom can effectively be reduced by functionally linking the muscles so that they mutually determine one another's states in a systematic fashion. But such linkage control would, in large part, be internal to the set of muscles. Such functional linkages, that render

an aggregate of relatively independent muscles into a single autonomous unit, may be conceived of as equations-of-constraint written, as it were, on the ascending and descending neural pathways.

To identify some important features of this latter system, let us compare it with closed-loop control in relation to the problem of uttering a vowel under conditions of efferent and afferent interference. In the closed-loop perspective, to produce a given vowel is to specify a particular spatial target as referent. In the coordinative structure perspective just outlined, to produce vowels is to organize the articulators into a single, autonomous system according to a particular equation (or set of equations) of constraint; and, to produce a given vowel is (perhaps) to parameterize that system in a particular way (cf. Fowler, 1977).

Suppose that a speaker impeded by a bite block is requested to utter a given vowel. The model-referenced version of closed-loop control assumes that the condition of the speech apparatus is sensed and motor commands together with sensory feedback are simulated to determine what needs to be done given these conditions. The coordinative structure perspective simply notes that if some parts of the system are 'frozen' the other parts will, by virtue of the equation(s) of constraint, automatically assume values tailored to that of the frozen part and appropriate for producing the vowel.

Suppose now a speaker is interfered with not by a bite block but by anesthetization of parts of the speech apparatus and, as before, is requested to utter a given vowel. In this situation model-referenced control must suffer to the extent that sensory information about initial conditions is not available. In short, anesthetization should impair vowel production considerably more than a bite block restriction. From the coordinative structure perspective, however, anesthetization and bite block should be equivalent in that neither one alone should seriously perturb vowel production. For some members of a coordinative structure to be 'uninformed' about the states of other members is not important; as long as all members of the structure can vary, equilibration according to the equation(s) of constraint will occur and vowel production will be successful. However, we suspect that if some members cannot vary (due to a bite block) and their values are not

communicated within the system (due to anesthetization), then fulfilling the equation(s) of constraint will not be possible and successful vowel production would be seriously hindered. The experiment that follows is a preliminary appraisal of these notions. In it, both efferent and afferent variables were either interfered with directly or controlled indirectly during the production of several isolated vowels. Both acoustic and electromyographic measures were used to determine how speech performance is affected when the linkage among these variables is both partially and completely disrupted.

Method

Subjects were two adult male native speakers of American English, one phonetically trained (WE) and the other phonetically naive (SJ). The speech material consisted of the isolated vowels, /i,a,u/. Four separate articulatory variables were controlled directly and one was controlled indirectly. A bite block and an artificial palate were used to produce direct efferent interference, and anesthesia of the temporomandibular joint (TMJ) and oral mucosa were used to produce direct afferent interference. Two different bite blocks were used, one 23 mm long and the other 3 mm long. The longer bite block was used to fix jaw position for the close vowels /i/ and /u/, and the shorter bite block was used for the open vowel /a/. An acrylic artificial palate was constructed from upper mouth casts of both subjects. This prosthesis was approximately 10 mm thick at the midline, 3 mm thick along its edge, and 5 cc in volume. It extended from the posterior surface of the central incisors to approximately 8 mm anterior to the soft palate. Jaw position afference from the mechanoreceptors of the temporomandibular joint was eliminated by 2 ml of xylocaine injected directly into the joint capsules, bilaterally. Oral mucosa sensation was eliminated by spraying the entire oral cavity with a benzocaine solution.

The recording procedures were as follows: First, the subjects produced three triads of each vowel spontaneously, with the bite block, the artificial palate, and the bite block and artificial palate, in that order. Anesthesia was then applied in two steps. For one subject (WE), the joint was anesthetized first, while for the other subject the topical anesthesia was applied first. In each case, the entire vowel sequence was repeated after

each anesthetization. The experiments were run with the subjects seated in front of a microphone. For one subject (WE), electromyographic recordings from the genioglossus (tongue) and orbicularis oris (lip) muscles were obtained using conventional hooked wire techniques. All data were recorded on magnetic tape for later analysis.

Results

For both subjects, the effects of the experimental conditions were variable and evident only for /i/ and /u/; the formant frequencies of /a/ were virtually unaffected by either mechanical interference or anesthesia. Apparently, only pharyngeal cavity variables are relevant to /a/. For both /i/ and /u/, articulatory performance was unaffected by anesthesia alone, the artificial palate under all conditions of anesthesia, and the bite block under normal conditions and under incomplete anesthesia. Performance was affected, however, and dramatically so, when the bite block was introduced either alone or in combination with the artificial palate under complete anesthesia. These effects were substantial not only perceptually, but at the acoustic and muscle activity level as well. For example, first and second formant frequencies of the first spontaneous /i/ produced by subject WE were 275 and 2275 Hz. These values were approached for all experimental combinations except the TMJ + topical anesthesia + bite block and TMJ + topical anesthesia + bite block + artificial palate conditions where first and second formant frequencies shifted to 375 and 2050 Hz and 425 and 1600 Hz, respectively. Formant shifts were also evident for subject SJ, although to a slightly lesser degree. The EMG data dramatically illustrate these effects. Figure 1 shows the genioglossus EMG for the spontaneous bite block and TMJ + topical anesthesia + bite block conditions for the vowel /i/. The top trace shows the genioglossus muscle activity for /i/ produced spontaneously. The middle trace shows the corresponding EMG for the simple bite block condition. The increase in activity here is expected because the tongue has farther to move from a fixed-open position toward its target. Note also that the increase in activity is present at onset, before any online feedback mechanism would have time to generate an adjusted movement. The lower record corresponds to the TMJ + topical anesthesia + bite block condition. It shows virtually no activity. Absence of muscle

activity was the rule for all tokens within this condition as well as for the TMJ + topical anesthesia + bite block + artificial palate condition. Apparently, fixation of both the efferent and afferent variables resulted in an inability to produce any coordinated movement; hence, a neutral tongue position and a tendency toward schwa.

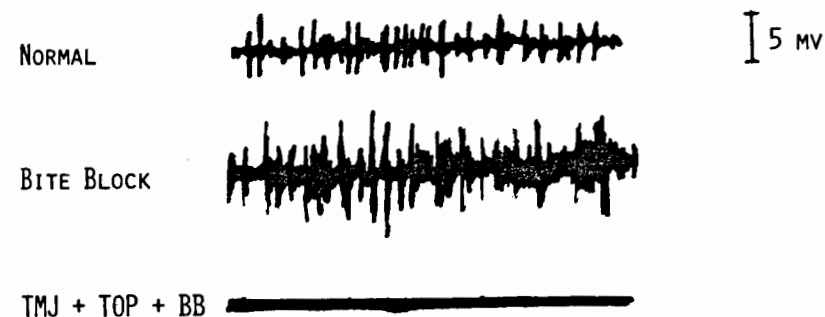


Figure 1

EMG activity for three experimental conditions.

This motor disorganization, however, was relatively short-lived. In each case learning took place and the normal vowel targets were reached after several trials. Table 1 shows the formant frequency values for /i/ produced under the most extreme experimental condition for each of nine token repetitions, for subject WE. Again, measurements were made at the time of the first glottal pulse. For even this extreme condition, complete acoustic compensation was attained by the sixth trial where vowel targets approached those of the spontaneously produced vowel.

Table 1

First and second formant frequencies for the vowel /i/ produced by subject WE under the most extreme experimental conditions (TMJ + topical anesthesia + bite block + artificial palate). Normal vowel target values are: F1 = 275 Hz, F2 = 2275 Hz.

	TRIALS								
	1	2	3	4	5	6	7	8	9
F1	425	500	475	325	325	275	300	300	275
F2	1600	1700	1900	2050	2150	2175	2225	2225	2250

Conclusions

The main finding of this experiment was that interference with either an efferent or afferent variable alone did not affect the production of isolated vowels; however, simultaneous interference with both efferent and afferent variables seriously altered vowel production. It is our view that these findings demonstrate both the necessity of a generative approach to speech production modeling and the utility of a coordinative structure mechanism for the control of speech movements. First, the experimental conditions produced novel physical and sensory situations that were met with immediate and successful articulatory responses. An open-loop model based on stored experiences cannot explain the success of these responses. Second, from the coordinative structure perspective, the finding that afferent interference does not affect vowel production unless an efferent variable is frozen is consistent with the view that these muscles are functionally linked across efference and afference in such a way that control can be taken over by either system when the other is fixed.

References

- Fowler, C.A. (1977): "Timing control in speech production", Indiana University Linguistics Club, Bloomington, Indiana.
- Fowler, C.A. and M.T. Turvey (1978): "Skill acquisition: An event approach with special reference to searching for the optimum of a function of several variables", to appear in Information Processing in Motor Control and Learning, G. Stelmach (ed.), New York: Academic Press (in press).
- Fowler, C.A., P. Rubin, R.E. Remez and M.T. Turvey (1978): "Implications for speech production of a general theory of action", Language Production, B. Butterworth (ed.), New York: Academic Press.
- Lindblom, B., J. Lubker and T. Gay (1978): "Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation", JPh (in press).
- MacNeilage, P.F. (1970): "Motor control of serial ordering of speech", Psychological Review 77, 182-196.
- Turvey, M.T., R. Shaw and W. Mace (1978): "Issues in the theory of action: Degrees of freedom, coordinative structures and coalitions", in Attention and Performance VII, ed. by J. Requin, Hillsdale, N.J.: Erlbaum.

A CORRELATION ANALYSIS OF EMG ACTIVITY AND THE MOVEMENT OF
SELECTED SPEECH ORGANS

Hajime Hirose, Research Institute of Logopedics and Phoniatics,
Faculty of Medicine, University of Tokyo, Tokyo, Japan

Introduction

In the study of the dynamic aspects of speech production, it is ultimately necessary to investigate the pattern of motor control signals from the central nervous system and the dynamic characteristics of the speech organ(s) which act(s) in response to the control signals. Although the pattern of the motor control signal has usually been observed in the form of electromyographic (EMG) potentials, the quantitative analysis of the relationship between EMG activity and articulatory movement has remained difficult. Cinefluorographic observation combined with simultaneous recording of EMG signals has been considered to be most satisfactory, but the acquisition of necessary information is generally restricted due to the dosage problem.

The introduction of the x-ray microbeam system to speech research (Kiritani, Itoh and Fujimura, 1975) solved the dosage problem to a large extent and, at the same time, proved useful for reducing the time required for data analysis. Figure 1 shows a data collection and analysis system in use at the Research Institute of Logopedics and Phoniatics, Faculty of Medicine, University of Tokyo.

The present study is an attempt to analyze the dynamic characteristics of the movements of selected speech organs recorded by means of our x-ray microbeam system simultaneously with EMG recordings of the activity of the related articulatory muscles during speech. In the present paper, the preliminary results of an analysis of velar movement will be presented as an example, in reference to the pattern of the EMG activity of the levator palatini muscle. The articulatory movement of the velum is known to be controlled almost solely by the activation and suppression of the levator palatini. Velar movement is relatively independent of the movement of the other articulators and, thus, relationship between the displacement of the velum and the EMG patterns of the levator palatini can be considered to be relatively straightforward.

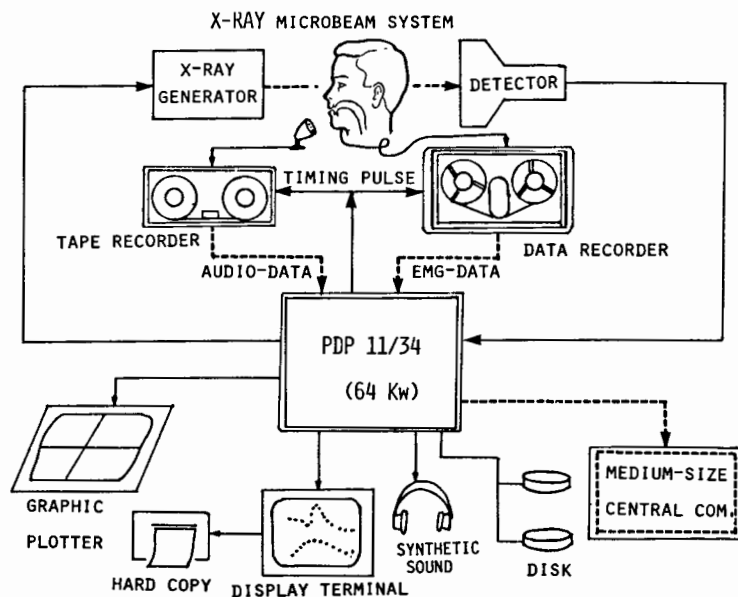


Figure 1 The Articulatory data collection and analysis system at the RILP, University of Tokyo

Data recording

An adult male speaker of the Tokyo dialect served as the subject of the present study. The subject read a list of test words either in isolation or embedded in appropriate frame sentences.

For recording the articulatory movements, the movement of lead pellets attached to the pertinent articulators was tracked and recorded by means of the computer-controlled x-ray microbeam system. For recording velar movement, a strip of thin plastic film with a lead pellet attached to its end was passed through a nostril, placing the pellet on the nasal surface of the velum. The pellet movement was recorded with a frame rate of approximately 190 frames/sec.

Conventional hooked-wire electrodes were inserted into the levator palatini in this particular case. The EMG signals were recorded on an FM data recorder together with the speech signals

and the timing pulses which were generated from the computer for each frame of the x-ray tracking. The EMG signals were then digitized through an A/D converter with a sampling rate of 8 kHz. Absolute values were taken and integrated over 5.83 msec, the value of which corresponds to the interval between successive timing pulses.

Data analysis

The Y-coordinate of the pellet on the velum was selected to represent the movement of the velum, and the relationship between the time function of the coordinate value and the EMG signal was examined. EMG activity is associated with the generation of muscle force and, therefore, can be related to such variables as the displacement, velocity and acceleration of the movement of the pertinent articulator. Thus, the present analysis aimed at obtaining a quantitative estimation of the relationship between these variables and the EMG signal.

It was assumed that the EMG activity of the levator palatini at a given instant could be expressed as the sum of the three components dependent on the displacement (y), velocity (\dot{y}) and acceleration (\ddot{y}) of the movement of the velum. Thus, an estimated EMG signal at a given time can be given as in equation (1).

$$\hat{E}_i = c_0 + c_1 y_i + c_2 \dot{y}_i + c_3 \ddot{y}_i \quad \text{--- (1)}$$

In this equation, the subscript i denotes the i -th time sample. The above equation indicates that velar movement is realized as the response of a linear second order system to the EMG signal which is given as input. The coefficients which give the best approximation were estimated by the least square error method. That is, for every time sample of EMG signal E_i , the estimate \hat{E}_i in equation (1) was formed by using the coordinate values obtained by x-ray tracking. The coefficients, c_{0-3} were determined by minimizing the value of error (E_{rr}) in equation (2).

$$E_{rr} = \sum_i (E_i - \hat{E}_i)^2 \quad \text{----- (2)}$$

In the above procedure, it was necessary to introduce a temporal smoothing of the observed coordinate value, since, without smoothing, the noise components in the calculated velocity and

acceleration were so dominant that virtually no effective correlation could be observed between these variables and the EMG signals. In order to reduce the noise effect, the temporal variation of the coordinate value within a short time window was approximated by the parabolic function of time. In the data sets obtained in the present study, it was found that the error was minimum for a time window of about 30 frames. Thus, in the present analysis, the values calculated for this time window width were considered to be the best estimates of the coefficients.

Results and discussion

The characteristic constants of the linear second order system were calculated from the estimated values of the coefficients in equation (1). The value of the damping factor was found to be close to 1, which implied that the second order system is critically damped. The characteristic time constant was approximately 80 msec, regardless of the difference in speaking rate.

Figure 2 shows examples of the x-ray and EMG data obtained. These curves correspond to the three different types of test words, /bemeē/, /beN'ee/ and /beNmee/, each of which was embedded in a frame "sorewa _____ desu" (that is _____). In the test words, /N/ represents the syllable final nasal element in Japanese. The sequence of nasal segments /Nm/ is generally uttered as a geminate nasal consonant. For each test word, the curve at the top shows the audio signal, the second and third curves are the temporal patterns of the Y-coordinate value for the lower lip and the velum, respectively. The bottom curve shows the integrated raw EMG, with the estimated EMG curve calculated by using equation (1) superimposed.

It can be seen that in /bemeē/ the velum lowering for /m/ starts immediately after the oral release of the initial /b/ and continues until the release of /m/. Velar elevation then begins, the speed of which appears to be slower than that of lowering, and, as a result, the temporal pattern of velar movement is asymmetrical for the production of /m/. In /beN'ee/, the velum lowering for /N/ continues longer, and the velar displacement is larger than for /m/ so that the temporal pattern appears to be symmetrical. In /beNmee/, the level of the maximum velum lowering for /Nm/ is higher than for /N/ in /beN'ee/, although the duration of nasaliza-

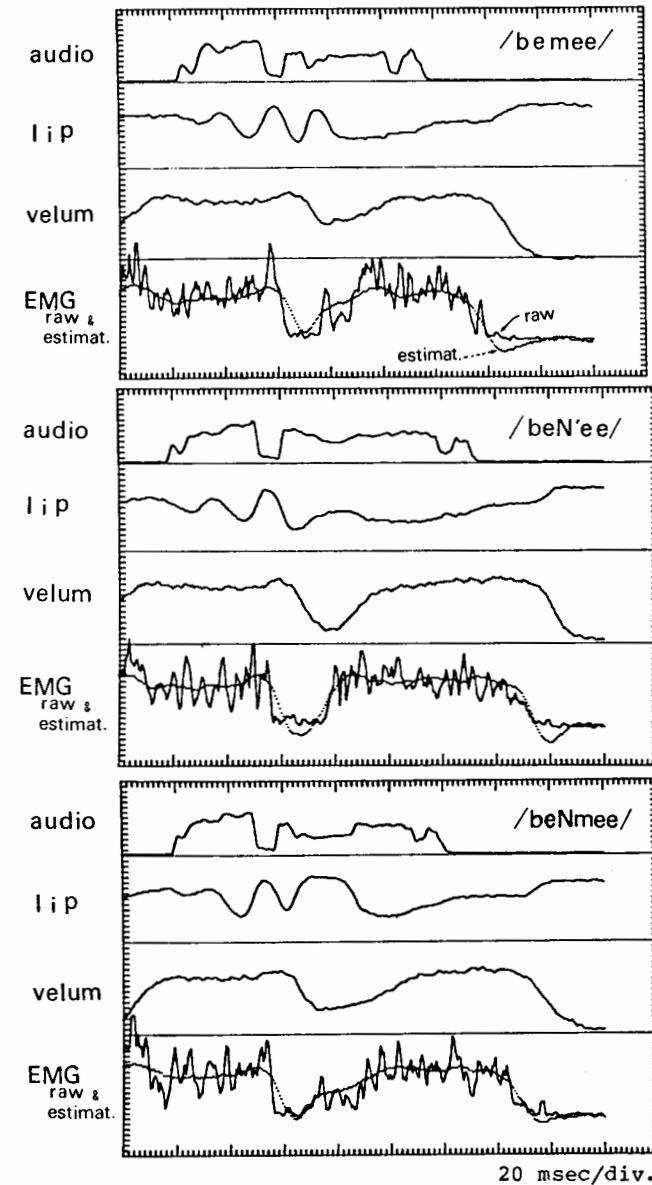


Figure 2 The temporal patterns of the Y-coordinate value for the lower lip and the velum, compared with the audio signal (top) and the raw and the estimated EMG curves (bottom).

tion is longer. In this case, after reaching the level of maximum lowering, the velum appears to stay at, or to ascend very gradually from, that level and, thereafter, it ascends with a speed similar to that for /m/ in /bemeē/.

Comparing the patterns of the raw EMG activity with those of the velar movement, it appears that the pattern of levator activity for /N/ in /beN'ee/ is characterized by a step-like EMG suppression, the level of which is the same as that for the resting position of the velum after the cessation of the utterance. In other words, the movement of the velum for /N/ can be regarded as a smoothed response of the velum as a second order system to the step-like control signal to the velum. In /beNmee/, on the other hand, a rapid suppression of levator activity is followed by a short period of an intermediate level of EMG activity. Although the EMG level of the initial part of the suppression for /Nm/ is apparently the same as that for /N/ in /beN'ee/, the velum does not show an extreme lowering but stays at a somewhat higher position. Thus, it appears that the period of intermediate level EMG activity mentioned above is responsible for the characteristic pattern of velum movement for /Nm/. The duration of the suppression of the levator activity for /m/ in /bemeē/ is relatively short compared to the estimated value of the time constant of the second order system, and, therefore, the pattern of the velar movement for /m/ can be taken as a ballistic impulse response to the EMG signal. However, as seen in Figure 2, a short re-activation of the levator palatini reaching the intermediate level is observed in some utterance samples of /bemeē/. Thus, for /m/, a question still remains as to whether the apparent re-activation of the levator palatini may necessarily result in the gradual ascent of the velum after the initial lowering.

The characteristic pattern of each of the three utterance types is also demonstrated in the estimated EMG curve. In particular, the estimated curve for /Nm/ is characterized by the fact that when the EMG activity increases after the negative peak of suppression, the rate of increase is temporarily depressed before it reaches the maximum level, and, as a result, there is a plateau in the estimated curve. This result would indicate that, as far as the second order linear relationship between the levator EMG and velar movement is concerned, the patterns of the velar movement

for /Nm/ can not be accounted for by a constant increase in the EMG activity after suppression. Rather, it can be assumed that an intermediate stage of EMG control is necessary during the phase of re-activation.

It has been reported that there are characteristic differences in the temporal patterns of velar movements for these three utterance types (Ushijima and Hirose, 1974; Fujimura, Miller and Kiritani, 1976). The result of the present study seems to confirm this result. It also suggests that these differences are based on the different patterns of motor control to the velum.

For a better understanding of the dynamic aspects of speech production, further attempts should be made to accomplish the quantitative analysis of the relationship between EMG signals and articulatory movements. A preliminary analysis of jaw movements in reference to the pattern of the EMG activity of the related muscles is now in progress.

References

- Fujimura, O., J.E. Miller and S. Kiritani (1976); "Syllable final nasal element in English - An x-ray microbeam study of velum height", 92nd Meeting of ASA, S 65.
- Kiritani, S., K. Itoh and O. Fujimura (1975); "Tongue-pellet tracking by a computer-controlled x-ray microbeam system", JASA 57, 1516-1520.
- Ushijima, T. and H. Hirose (1974); "Electromyographic study of the velum during speech". JPh 2, 315-326.

Acknowledgement

This research was supported in part by Grant in Aid for Scientific Research, Ministry of Education (No.349008).

ON THE USE OF OROSENSORY FEEDBACK: AN INTERPRETATION OF
COMPENSATORY ARTICULATION EXPERIMENTS

Joseph S. Perkell, Research Laboratory of Electronics,
Massachusetts Institute of Technology, Cambridge, Mass. 02139,
U.S.A.

A number of experiments have been performed on "compensatory articulation" with the aim of understanding more about speech motor programming. Several of these experiments have used bite blocks to constrain the mandible in abnormally open (or closed) positions while the subjects produced steady state vowels (/i/, /a/, and /u/) (cf. Lindblom, 1971; Lindblom, et al., 1977; Lindblom et al., in press; Gay and Turvey, these proceedings). The resulting formant patterns were measured at the first glottal pulse to avoid any possible effects of auditory feedback (which was not masked out). It was found that vowels produced with significantly abnormal jaw openings (i.e. 22-25 mm open for /i/) were essentially the same in quality as those produced normally by the same subjects. However when bite blocks were used in conjunction with oral topical anesthesia (Lindblom, et al., 1977) or with a combination of oral topical anesthesia and anesthesia of the temporomandibular joint (Gay and Turvey, these proceedings), subjects needed several attempts to produce appropriate vocal tract configurations and sound outputs. In the latter experiment, the application of oral topical anesthesia alone was not enough to impair subjects' ability to produce vowels appropriately.

Lindblom and his co-workers interpret their findings as support for the following view of the role of orosensory feedback. Tactile information from the labial and oral mucosa can be utilized in the motor programming of speech. Vowel "targets" may be encoded as [oro]sensory goals which reflect a neuro-physiological encoding of area functions. These goals serve as a basis for the elaboration of motor commands by structures which "can generate appropriately revised motor commands on the basis of the feedback positional information available before onset of phonation" (Lindblom, et al., in press).

These results and their interpretations must be viewed with caution for a number of reasons. For example, a generous application of topical anesthesia to the oral and pharyngeal cavities can have a distracting effect on the subject (Lindblom,

personal communication). Perhaps more importantly, a steady-state paradigm which allows the subject time to "organize" his response before presenting it may reflect functions which are not part of normal dynamic speech motor processes (cf. Leanderson and Persson, 1972; Abbs and Eilenberg, 1976). Nevertheless, the results are provocative enough to warrant further examination, particularly in light of a recent experiment on arm movements and another experiment on compensatory articulation.

Polit and Bizzi (1978) have performed an experiment in which 3 adult monkeys were trained to point to a target light with the forearm and hold the position for about 1 second in order to obtain a reward. The monkey could not see its forearm which was fixed to an apparatus that permitted only flexion and extension about the elbow in the horizontal plane. Performance was tested before and after a dorsal root section which eliminated somatosensory feedback from both upper limbs. In both intact and deafferented animals, the arm was unexpectedly displaced within the reaction time of the monkey, and in both cases the displacement of the initial arm position did not affect the attainment of the intended final steady-state position. These results suggested to the authors that a central program specified an equilibrium point corresponding to the interaction of agonist and antagonist muscles. A change in the equilibrium point leads to movement and attainment of a new posture.¹ However, it was also found that when the spatial relationship between the animal's arm and body was changed, the pointing response of the deafferented monkeys was inaccurate, and remained so even when visual feedback was allowed. In contrast, the intact monkeys were able to compensate within a few tries to the new position without visual feedback. This finding suggested that one major function of afferent feedback is in the adaptive modification of learned motor programs (Polit and Bizzi, 1978).

Following these authors' interpretation of their results, we might consider that in establishing the central program for the performance of the motor task (i.e., learning the task), the monkeys were incorporating a subconscious "knowledge" of the relationships between the target points with respect to the

(1) The existence of additional processes related to the dynamic aspects of movements is acknowledged, but not treated by Polit and Bizzi (1978).

apparatus and the muscular settings which would result in correct pointing. In doing so, the monkeys were calibrating the biomechanical properties of the system with respect to a particular frame of reference (i.e. orientation in space in relation to the body) with the use of somatosensory feedback from the system. When the frame of reference changed, only the monkeys with intact somatosensory pathways were able to "recalibrate" the central program to the new frame of reference.

This line of reasoning and the interpretations of Lindblom and his colleagues lead us to a possible, slightly more specific explanation of the compensatory articulation results. In the case of steady-state vowel productions, the frame of reference is defined as the configuration of the dorsal walls of the vocal tract and the position of the mandible. The target (or goal) consists of a vocal-tract area function as sensed by a complex pattern of sensory feedback from the vocal tract. Normally, to produce a steady-state configuration, the control mechanism has a choice of: 1) using a pattern of peripheral feedback to compare with one that has been learned in association with a particular area function and vowel quality, or 2) using a set of equilibrium levels of muscle excitations. These muscle excitation levels can be stored or computed on the basis of an overlearned knowledge of the vocal-tract geometry and biomechanical properties.

Now let us consider the three possible combinations of the use of anesthesia and/or bite blocks. With only (complete) anesthesia, the controller uses option 2. In other words, with a frame of reference which is assumed to be normal, the controller is still capable of specifying equilibrium muscle excitations which it "knows" will produce the correct area function. On the other hand with only the bite block, the controller uses option 1. The appropriate area function is produced by comparing peripheral feedback with the "known" pattern. With anesthesia and the bite block, neither option is available. The frame of reference has been changed. The absence of feedback about the new frame of reference precludes an a priori recomputation of appropriate equilibrium muscle excitations, and the absence of tactile feedback precludes a direct comparison with the known pattern. This last statement is reinforced by the results of Gay and Turvey in which only combined anesthetization

of the oral mucosa and the temporomandibular joint (along with the bite block) rendered the subject incapable of producing the vowel correctly on the first try. The loss of joint sensation would eliminate the feedback about the frame of reference, needed for a recalibration of the central program, and the loss of sensation from the oral mucosa would preclude using such feedback directly in an error-minimizing feedback loop.²

The hypothetical use of afferent information to keep the controller informed about the frame of reference would be equivalent to the function proposed by Polit and Bizzi (1978) in the adaptive modification of learned motor programs. Presumably, the predictive simulation mechanism proposed by Lindblom, et al., (in press) also needs to use feedback in a similar way. It has been suggested by numerous investigators that learning a motor activity consists in part of substituting central programming for the use of peripheral feedback. This use of central patterning presumably incorporates an ability to adjust the parameters of the central program to account for changes in the frame of reference. In the case of speech, such changes correspond to speaking with a pipe clenched between the teeth, with the head tilted to one side, or resting ones' chin in his or her hand.

Gay and Turvey also found that /i/ and /u/ productions are affected by the combination of joint and topical anesthesia with a bite block while /a/ is not. This difference might be explained by their report that the topical anesthesia was applied to the oral cavity, where the acoustically most critical points of maximal constriction for /i/ and /u/ are located. (A given change in the dorsal-ventral location of the tongue surface will have a proportionally larger effect on the vocal-tract cross-sectional area at the point of maximum constriction than at other locations where the cross-sectional area is greater.) If the anesthesia did not exert a strong effect at the point of maximal constriction for /a/, we might expect it to be produced normally, with the use of feedback which is less consciously obvious but still available from that region. The importance of the pattern of contact at

(2) The fact that only topical anesthesia in combination with bite block was sufficient to impair vowel production in the subject of Lindblom, et al. (1977) might be due to individual differences or differences in the extent and depth of topical anesthesia.

the point of maximal constriction for the vowel is further suggested in lateral cineradiographic tracings (Netsell, et al., 1978) of normal and compensatory productions of the vowel /i/ for 3 subjects. For each subject the normal and compensatory dorsal tongue contours show considerable overlap and the overlap is most pronounced at or near the point of maximal constriction.

This brief analysis greatly oversimplifies the issues in a number of respects. It relies on a small amount of data. It overlooks the significant differences between deafferentation and the application of anesthesia as well as the unnatural nature of both experimental paradigms. And as we have mentioned, it deals with a steady-state task which may be quite different from anything actually found in speech. For these reasons, we must be very tentative in extending our interpretations to cover normal articulatory movements. However, on the basis of considering a number of additional aspects of speech production and the control of movement (see Perkell, 1979a), it is possible to offer the following speculations on the use of orosensory feedback.

1) Orosensory feedback may play a role in determining the nature of some distinctive features. It is possible that certain well-defined patterns of orosensory feedback (such as contact of the tongue with maxillary structures) facilitate the production of sounds which have distinctive acoustic and auditory-perceptual correlates (see also, Stevens and Perkell, 1977, Perkell, 1979b). Such patterns of orosensory feedback could be the speech production correlates of distinctive features. Specifications of utterances in the form of feature-related complexes of orosensory goals might serve as a basis for the production of articulatory movements. Thus, orosensory feedback on a long-term basis might be necessary for the establishment and maintenance of a sub-conscious "knowledge" of the orosensory correlates of the features. This "knowledge" could be used directly as suggested by the bite block results or indirectly in the establishment and maintenance of central programs.

2) As suggested by the discussion in this paper, orosensory feedback might be important in informing any central programming mechanism about the overall state of the system or frame of reference. The use of feedback to make adjustments for changes in the frame of reference could cover a time span corresponding

to several movements in a sequence (Polit and Bizzi, 1978). In more general terms, perceptual feedback "regarding the position or movement trajectories of one or more articulators could be used for preprogramming movements several hundred milliseconds into the future." (Larson, personal communication).

3) This paper has implied that central programming plays a major role in the production of articulatory movements. Much of the experimental and theoretical work on other forms of movement suggests that central programming along with internal feedback (feedback entirely internal to the central nervous system) is used for the moment-to-moment (context-dependent) programming of rapid movement sequences. While this is most likely the case for "learned" or "skilled" motor behavior such as speech production (cf. Lindblom, et al., in press), we must keep in mind that vocal tract motor control mechanisms may conceivably have capabilities that other systems do not have (cf. Folkins and Abbs, 1975, 1976, McClean, et al., in press). Thus it is possible that orosensory (peripheral) feedback from the vocal tract is used on a moment-to-moment basis to assist in the programming of articulatory movements in ways that have not been demonstrated for other types of movement.

Ideas such as these are closely related to questions about the nature of fundamental units which underlie the programming of speech production (cf. MacNeilage, 1970). Thus, the difficulty of testing such hypotheses should not stand in the way of exploring them further. The striking similarities between the compensatory articulation and arm movement results discussed above suggest that we may learn increasingly more about speech by continuing to follow future work on analogous types of movement.³

References

Abbs, J.H. and G.R. Eilenberg (1976): "Peripheral mechanisms of speech motor control", in Contemporary Issues of Experimental Phonetics, 139-168, N.J. Lass (ed.), New York: Academic Press.

(3) I am very grateful to Profs. Emilio Bizzi of M.I.T. and Stephanie Shattuck-Hufnagel of Cornell University for their comments on parts of this manuscript. I also thank Dr. Charles Larson of the University of Washington for his very helpful comments on a previous manuscript. This work was supported by National Institutes of Health Grant NS04332.

- Folkins, J.W. and J.H. Abbs (1975): "Lip and jaw motor control during speech: Responses to resistive loading of the jaw", JSHR 18, 1, 207.
- Folkins, J.W. and J.H. Abbs (1976): "Additional observations on responses to resistive loading of the jaw", JSHR 19, 820-821.
- Leanderson, R. and A. Persson (1972): "The effect of trigeminal nerve block on the articulatory EMG activity of facial muscles", Acta-Oto-Laryngologica 74, 271-278.
- Lindblom, B.E.F. (1971): "Neurophysiological representation of speech sounds", Publication 7, Papers from the Institute of Linguistics, University of Stockholm.
- Lindblom, B., R. McAllister, and J. Lubker (1977): "Compensatory articulation and the modeling of normal speech production behavior", Paper presented at the Symposium on Articulatory Modeling, Grenoble, France, July 11-12.
- Lindblom, B., J. Lubker and, T. Gay (in press): "Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation", JPh.
- MacNeilage, P.F. (1970): "The motor control of serial ordering in speech", Psychological Review 77, 182-196.
- McClellan, M.D., J.W. Folkins, and C.W. Larson (in press): "The role of the perioral reflex in lip motor control for speech", Brain and Language.
- Netsell, R., R. Kent, and J. Abbs (1978): "Adjustments of the tongue and lips to fixed jaw positions during speech: a preliminary report", Paper presented at the Conference on Speech Motor Control, University of Wisconsin, Madison, June 2-3.
- Perkell, J.S. (1979a): "Phonetic features and the physiology of speech production", in Language Production, B. Butterworth (ed.), New York: Academic Press.
- Perkell, J.S. (1979b): "On the nature of distinctive features: implications of a preliminary vowel production study", in Frontiers of Speech Communication Research, B. Lindblom and S. Ohman (eds.), London: Academic Press.
- Polit, A. and E. Bizzi (1978): "Processes controlling arm movements in monkeys", Science 201, 1235.
- Stevens, K.N. and J.S. Perkell (1977): "Speech physiology and phonetic features", in Dynamic Aspects of Speech Production, 323-341, M. Sawashima and F.S. Cooper (eds.), University of Tokyo Press.

MOTOR UNIT DISCHARGE PATTERNS DURING SPEECH: TEMPORAL REORGANIZATION DUE TO COARTICULATORY AND PROSODIC EVENTS

Harvey M. Sussman, Department of Linguistics, University of Texas, Austin, Texas, U.S.A.

The functional unit of muscle contraction, and hence movement control, is the motor unit. A motor unit consists of an alpha motoneuron, located, in the case of speech muscles, within a motoneuron pool in the brainstem, and the single muscle fibers innervated by the axonal branches of that motoneuron. For the past few years my colleagues and I have been interested in describing how motor unit discharge properties change to meet the demands of rapid tension development characteristic of speech production. Observation of motor unit discharge activity represents the closest look at the encoding operations of the central nervous system as the peripherally recorded muscle action potentials (MAPs) of motor units (MUs) stand in a 1:1 relationship to discharges of centrally located alpha motoneurons.

Data will be restricted to MU events within the anterior belly of digastric (ABD), a muscle involved in lowering the jaw during speech. Specially constructed intramuscular wire electrodes, designed to facilitate recording at high force levels, were used in all studies. All measurements were performed on a digitized oscilloscopic display (maximum resolution = 0.1 msec) utilizing computer software routines written for temporal and statistical analyses of motor unit/articulator events during ongoing speech (complete details of recording and measurement procedures can be obtained from Sussman et al, 1977).

Recruitment Order

A current view explaining activation of individual MUs is the Size Principle (Henneman et al, 1965). Briefly stated, the

Size Principle holds that MUs are activated according to motoneuron size, with the smallest neurons (having the lowest excitability thresholds) discharging first, followed by successively larger motoneurons. Evidence supporting this view has been extensively gathered, but primarily from animal experimentation. Recruitment order of MUs active during human isotonic movements, such as speech, have not received much attention to date. Our data overwhelmingly supports the Size Principle. Recruitment order of ABD MUs was observed to be fixed and based on size (as determined by peak-to-peak amplitudes of MAPs). The consistency of MU activation for jaw lowering represents an invariant aspect of the encoding program for speech.

Data on recruitment order can also be related to various aspects of articulatory dynamics. Both jaw displacement and velocity were found to be positively related to the number of MUs active (Sussman et al, 1977). In addition, the initial interspike interval (ISI) of the third recruited MU has consistently been shown to be linearly related to both jaw displacement and velocity. During jaw lowering for the initial vowel in /aepae/ tokens, it was found that jaw displacement and velocity increased as the initial ISI decreased. Correlation coefficients ranged from $-.44$ to $-.67$ and were significant beyond the $p < .05$ level for all utterances examined. The relationship between discharge rate of a MU and some aspect of articulatory dynamics was only found for the larger and later recruited MUs (specifically the third MU recruited). The smaller first and second MUs recruited did not exhibit a straightforward relationship between its initial firing rate and jaw movement. Since the larger and later recruited MUs add a proportionately larger contribution to overall tension development (i.e. larger MUs have higher twitch tension levels)

compared to initially active, smaller MUs, it is not surprising to notice movement variables being influenced by discharge characteristics of the larger MUs only.

Temporal Reorganization: Coarticulatory Influences

The temporal interval separating activation of the first recruited MU and the initiation of jaw lowering for an open vowel such as /ae/ can be a valuable dependent variable in providing a glimpse at the time program applied to the events of speech motor control. Such an analysis was made for 64 utterances of /aepae/ with separate measurements taken for initial vowel lowering and final vowel lowering. The results are schematically illustrated in Figure 1. For all utterances the first discharge of the first recruited MU occurred approximately 40 msec after the jaw began to lower for V1, and approximately 28 msec prior to jaw lowering for V2 opening gestures. These consistent differences (across three subjects) can be related to the differences in peripheral biomechanics existing at the moment of jaw lowering for the initial preconsonantal vowel versus the final postconsonantal vowel. It is well known that the jaw exhibits anticipatory coarticulation for an open vowel in final position of VCV tokens (Sussman et al, 1973) Thus, the jaw is lowering for the postconsonantal vowel from a position that is considerably lower than the jaw position preceding the initial preconsonantal vowel. Abbs and Eilenberg (1976) have shown that the mechanical advantage of ABD in exerting a lowering force on the mandible decreases the more the jaw is lowered. This reduction of mechanical advantage represents a diminution of the effective muscle force of ADB to bring about additional lowering for the postconsonantal vowel /ae/. The earlier activation of the initially recruited MU for the final postconsonantal vowel as compared to the initial preconsonantal

vowel may reflect a temporal adjustment of the motor time program needed to partially offset the less favorable mechanical advantage of the jaw during this time. It is consistent with this hypothesis that there was a highly significant positive correlation ($r = .74$, $p < .01$) between jaw position during the medial consonant and temporal onset of MU I. Thus, the lower the jaw immediately prior to subsequent lowering for the postconsonantal vowel, the earlier did MU I activity begin for jaw lowering for V2. This example provides the first illustration of temporal reorganization, on the cellular level, to a behavioral and biomechanical aspect of the encoding program for speech.

Temporal Reorganization: Stress

Previous studies investigating articulatory reorganization due to stress have shown that higher levels of integrated EMG signals, higher rates of articulator movement, and closer approximations of intended target positions accompany high stress conditions. Until recently, there have been no descriptions of temporal change in motor unit discharge patterns due to the prosodic application of syllable stress.

A subject repeated /aepae/ twenty times with equal "moderate" stress on each syllable and twenty times with heavy stress on the second syllable. The first three recruited MUs were examined for both stress conditions. Figure 2 shows recruitment latencies separating the initial discharges of the first three recruited MUs in conjunction with the temporal onset of jaw lowering for the second syllable for both /aepae/ and /aepæe/ tokens. A temporal starting point, $t = 0$, was taken to be the onset of MU I's initial spike. For stressed utterances there was a consistent shortening of the intervals separating successively recruited MUs and a shorter latency between MU I's initial discharge

and the moment of jaw lowering for /ae/. Table 1 gives data characterizing the temporal reorganization pattern in terms of means (in msec), standard deviations, and variability coefficients (SD/\bar{X}). Not only was the time program advanced for the stressed condition, but, in addition, there was a marked reduction in variability. The percent reduction in variability, calculated by comparing the unstressed and stressed variability coefficients, revealed a 60% reduction for the MU I - MU II interval, a 82% reduction for the MU I - MU III interval, and a 62% reduction for the MU I - jaw lowering interval.

Other changes in discharge characteristics accompanying stress were an increase in mean firing rate (impulses/sec) for each MU and a decrease in the mean initial interspike interval. This later parameter is indicative of a higher instantaneous discharge rate ($1/\text{initial ISI}$) for the stressed syllables. In addition to the higher instantaneous discharge rates for all three MUs, there was also a marked reduction in the variability of the initial ISI, with a progressively larger percent decrease in variability coefficients with recruitment order -- 21% reduction for MU I, 30% for MU II, and 42% for MU III. Thus, there was a "tighter" control over the onset times for the larger and later recruited MU.

These preliminary findings showing a sharp reduction in the variability of recruitment intervals (e.g. MU I-MU II), activation intervals (MU I-jaw lowering), and initial interspike intervals, add a new dimension to our understanding of articulatory movements underlying syllable stress. In addition to the connotations that go along with the familiar Ohman notion of "an instantaneous addition of a quantum of physiological energy" (Ohman, 1967, p. 33) for stressed productions, the encoding program for speech,

as observed in our data, suggests a more precise control of the timing of cellular events, as well as a more forceful execution of the peripheral dynamics.

Motor unit events and their systematic changes during various conditions of ongoing speech can be sensitive indicants of higher level linguistic conditions. Alterations in the temporal program underlying muscle and hence articulator activation can be observed at the level of the alphamotoneuron (at least indirectly that is).

References

- Abbs, J. H. and G. R. Eilenberg (1976): "Peripheral mechanisms of speech motor control," in Contemporary issues in experimental phonetics, N. J. Lass (ed.), 139-168, New York: Academic Press.
- Henneman, E., G. Somjen and D. O. Carpenter (1965): "Functional significance of cell size in spinal motoneurons," J. Neurophys. 28, 560-580.
- Ohman, S. (1967): "Word and sentence intonation: A quantitative model," STL-QPSR 2-3, 20-54.
- Sussman, H. M., P. F. MacNeilage, and R. J. Hanson (1973): "Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations," JSHR 16, 397-420.
- Sussman, H. M., P. F. MacNeilage, and R. K. Powers (1977): "Recruitment and discharge patterns of single motor units during speech production," JSHR 20, 613-630.

Table I: Means (in msec), standard deviations, and variability coefficients (SD/ \bar{X}) for various temporal intervals characterizing motor unit/articulatory events during unstressed and stressed tokens.

	MU I → MU II	MU I → MU III	MU I → Jaw Lowering
Unstressed	\bar{X} 31.9	40.9	53.8
	SD 23.2	41.3	53.3
	VC .7273	1.0098	.9907
Stressed	\bar{X} 23.7	26.6	33.4
	SD 6.8	4.9	12.8
	VC .2869	.1842	.3832

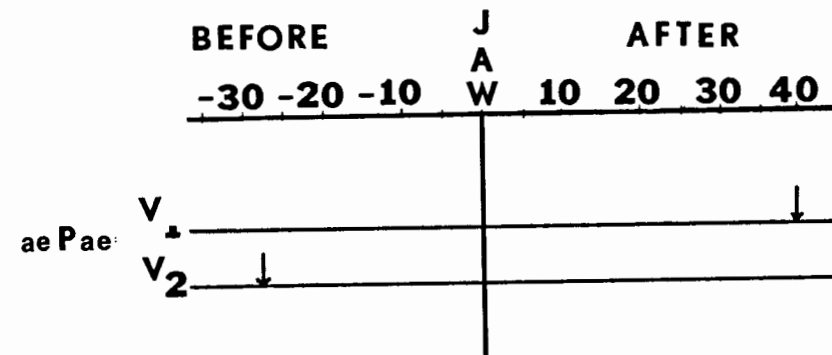


Figure 1: Temporal onset (msec) of initial discharge of the first recruited motor unit with respect to jaw lowering for initial (V1) and final (V2) vowel.

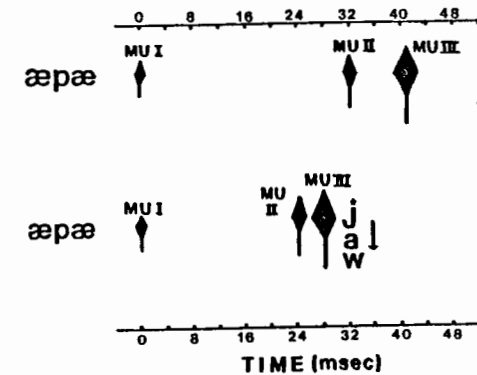


Figure 2: Recruitment latencies separating initial discharges of three recruited motor units during normal and stressed [æpæ]. Asterisks show onset of jaw lowering for open vowel of second syllable.

THE RELATION BETWEEN SENTENCE PROSODY AND WORD PROSODY

Summary of Moderator's Introduction

Eva Gårding, Swedish Research Council for Humanistic and Social Sciences and Phonetics Department, Lund University

The contributions to this symposium cover different kinds of prosodic systems and represent different methods of analysis and description. As a moderator I am taken aback by this variety but it may of course help us reach the goal set by the organizers: A pursuit of universal features in the relation between word prosody and sentence prosody. This requires a certain minimum of common terminology for the basic concepts and aspects of prosody. In my summaries and the points suggested for discussion I have used terms from the table below. Synonyms from other authors have been put within parentheses.

Word level	Sentence level	
	local effect	overall (general) effect
tone:	tone (term. junct.):	intonation:
contour tones	downglide	downdrift (declination)
level (static)		absence of downdrift
H (High)	upglide	
L (Low)		
accent (stress):	accent (stress)	
pitch accent		
?		
Citation form (ideal manifestation, ideal shape): a lexical item plus declarative sentence intonation and sentence accent		
Basic form: abstract representation or concrete form freed of other prosodic features		

SummariesThai. A. Abramson

The author is mainly interested in the question of whether the effects of sentence prosody are strong enough to weaken or destroy the lexical tones, the shapes of which have been derived from citation forms. Apart from perturbation from initial consonants there is perturbation from neighbouring forms. Also, in compounds one tone may be replaced by another (sandhi). In running

speech the tones are preserved but their shapes may undergo severe distortions. Sentence intonation can be marked locally by final particles and final sentence tone (terminal junctures), e.g. rising pitch for doubt, falling for finality, sustained for continuation and/or by overall effects. The author examines interspeaker frequency shifts in terminal junctures, expressed as percentages of voice range. Sentence accent is marked by lengthening, increase of amplitude and ideal tone shapes.

Some Southern Nigerian Languages. K. Williamson

The seven languages reported on have two basic tones, H and L. Two have distinctive downstep, H[↓]H versus HH. Tone rules express sequential modifications such as downdrift and coarticulation, glides between discrete tones and in one case replacement (cf sandhi in Thai). Sentence type can be marked locally (sentence tone) by an added L or H, or by replacement, or by overall intonational effects, e.g. downdrift or absence of downdrift, or by combinations of local and overall features. There is no consistent pitch signal for statement or question in these languages. Yet, statement is generally formed by downdrift of the basic tones, whereas for question some sentence-rule has to be added. Exclamations are uniform: a larger range by raising H's. Sentence accent is not mentioned.

Some American Indian Languages, etc. E.V. Pike

The author examines accent (stress) at word and sentence level in nine languages which use two or more contrastive tones. The tonal contrasts occur in both accented and unaccented syllables in eight of the nine languages. One language, Fasu, has tonal contrasts only on accented syllables. Apart from special allotones, such as raised H or lowered L, loudness and length (in consonant or vowel) mark the accented syllable. There are language-dependent rules for the distribution of word and sentence accent. Word accent is connected with the first, last or penultimate syllable of the stem. Sentence accent may fall on the last word accent or on any accented syllable, or it may have a separate manifestation, added to the final syllable. In some of the languages downdrift is reported for statement and downdrift + upglide for question. The prepause syllable has a special tone system for attitudes in two of the languages. Glottal stop, expressing finality, belongs in one of them, along with final-syllable upglide denoting politeness.

Swedish. G. Bruce

The author presents and modifies a model for Swedish intonation (Bruce and Gårding 1978). In this model the word accents (A1, A2) have been analysed as HL's with similar but differently timed Fo-correlates. There are dialectal differences but the HL occurs later in A2 than in A1 in all cases. Sentence accent is manifested as a wide pitch range with dialect-dependent distribution rules. In the main dialect areas a separate H is added after the word accent HL, in other dialects the wide range is obtained together with the accented syllable. Statement intonation is expressed as a progressing downdrift of H's and L's. The author argues that the Fo drop has a stepwise rather than a gradual downdrift. The downsteps cooccur with the accented syllables. Sentence intonation, then, has a systematic influence on Fo-values of H's and L's with higher values for earlier positions in the utterance.

Danish. N. Thorsen

A descriptive model of sentence intonation in Copenhagen Danish is presented. It does not take account of the word accents, stød (A1) and non-stød (A2). Sentence intonation is defined as a line described by the pitch of the accented syllables. An accented syllable and the following unaccented ones form a stress group in which the accented syllable is low in pitch and the unaccented syllables rise above and fall below the sentence-intonation line. This line is steeply falling for declarative sentences and level for unmarked questions. Sentence intonation has a systematic influence on the Fo course in the stress groups, in that the rise from stressed to unstressed syllable is larger in questions than in statements. The author rejects sentence accent but accepts emphatic or contrastive stress with manifestations common to many languages: a raising of pitch on the emphatic syllable at the expense of surrounding accents, i.e. shrinking of pre-emphatic and deletion of post-emphatic pitch movements in connection with accents.

Dutch. 't Hart and R. Collier

Each word has a lexical accent whose location can be predicted by rule. Under the influence of intonation it may be manifested as pitch movement. The authors study the interaction between intonation and accentuation. Two principles are discussed. According to the first, the overall pitch contour is obtained by adding

autonomous accentual pitch movements to autonomous pitch movements associated with sentence type, such as downdrift (declination) for statement and downdrift plus final upglide for question. This is rejected in favour of a second principle according to which the accents only determine the location of the pitch movements. Their nature (rise, fall, etc.) and order are determined by the chosen intonation pattern.

Czech. P. Janota

Janota reports on a series of tests with bisyllabic synthetic stimuli. When both syllables have the same pitch value, the first is judged as accented 85% of the time. A small increase or decrease of F_0 in the second syllable raises the number of accent votes for this syllable from 15 up to about 60%. With a larger change of F_0 , the number of such votes goes down. Responses to similar items, used to evaluate sentence intonation, show that precisely the stimuli with the large F_0 deviations are effective cues to intonation. Stimuli with a moderate or substantial increase of F_0 are judged as continuative statements and questions, respectively, and those with a decrease of F_0 as statements.

Suggested points for discussion

1. Universality of prosodic units

It may be fruitful to accept different degrees of universality, e.g. universal for similar function and similar acoustic correlates for all languages, and near-universal for corresponding conditions in almost all languages. As a third degree I suggest generality, requiring similar function, similar acoustic correlates and many languages.

Sentence intonation is a universal. This may be considered as a postulate. On the other hand, the contributions to this symposium and other communications show that statement intonation expressed as downdrift is merely a generality. The same holds for question intonation expressed as an absence of downdrift (Thorsen, Williamson).

Is sentence accent a universal? Is the lack of sentence accent in the description of some languages (Thorsen, Williamson) due to some possibilities in these languages to express deictic function at the sentence level in a non-prosodic way? Or is their description an artefact due to difference of analysis and tradition?

2. Principles for the analysis of the interaction between sentence prosody and word prosody. In recent work on Swedish intonation, basic forms of prosodic units have been isolated, after a phonological analysis, and rules for their combinations have been formulated (Bruce).

Other methods have also been mentioned. Is it possible to find a common framework applicable to all prosodic systems?

3. Universal features in the interaction between sentence prosody and word prosody.

't Hart and Collier demonstrate that for Dutch, sentence intonation is primary to word prosody. Is this a universal feature if we take primary in the following sense: In production sentence prosody precedes and sets the scale for word prosody. On the other hand, the degree to which word prosody interferes varies from very little as in Czech to something quite drastic as in Thai. Dutch seems to occupy an intermediary position and a 2-accent system like Swedish or Danish is closer to Thai.

Tone rules are a formal convenient way of expressing both co-articulation and the influence of sentence prosody on word prosody (see e.g. Williamson and her references). How universal are the tone rules?

4. Additional questions

Accent or Tone. In her description of various tonal languages Pike mentions one, Fasu, which has tonal contrasts (H, L) only on stressed syllables. What is the difference between a tonal language like Fasu and an accent language like Swedish in which the stressed syllables in some analyses (e.g. Malmberg) also have been represented by H versus L?

Accents and accents. Are there different physiological mechanisms behind different kinds of accent, e.g. the small pitch movements noted in Czech (Janota) and the larger ones typical of Germanic languages (Bruce, Thorsen, 't Hart and Collier). Or are they merely weak and strong manifestations of the same phenomenon?

Note

Two books have just come out which may be relevant for the discussion:

Fromkin, V. (ed.) (1978): Tone. A Linguistic Survey, Academic Press.

Greenberg, J. (ed.) (1978): Universals of Human Language. Volume 2. Phonology, Stanford University Press.

LEXICAL TONE AND SENTENCE PROSODY IN THAI

Arthur S. Abramson, University of Connecticut, Storrs, Connecticut, U.S.A. and Haskins Laboratories, New Haven, Connecticut, U.S.A.

Background

In a true tone language, one in which, in principle, every syllable in the morpheme stock bears a distinctive tonal phoneme, the tones are characterized primarily by fundamental-frequency levels and contours. Since we also describe intonation mainly in terms of the fundamental frequency (F_0) of the voice, there seems to be a paradox involved in examining the relations between sentence prosody and word prosody in a tone language. As in other languages so also in tone languages is there the possibility of expressing attitudes or indicating certain aspects of syntactic structure by means of sentence intonation. The question arises as to whether the effects of this sentence intonation are strong enough to weaken or even destroy the phonetic integrity of lexical tones.

The citation form of a monosyllabic word may be viewed as bearing the ideal manifestation of a tone. Of course, except for the occasional one-word sentence, these ideal forms do not often occur in running speech, yet children in the culture probably learn new words this way, and so do adults in a foreign-language class. Once we have two or more tone-bearing syllables strung together, we expect perturbations through coarticulation. The final physical shaping of a tone is provided by the intonation on the utterance (Pike, 1948, 18-19).

The Tones of Thai

The ideal shapes of the tones of Standard Thai (Siamese) have been described elsewhere (Abramson, 1962; Erickson, 1974). It is useful to divide the five distinctive tones of the language into the "dynamic" class, comprising the falling and rising tones, and the "static" class, comprising the high, mid and low tones. The dynamic tones show rapid movements of F_0 , while the static tones show rather slow movements which sometimes approximate levels. Of the three static tones it is the mid tone that is most likely to appear occasionally as a level. The high tone is more likely to be seen as a rise high in the voice range in contrast with the low rise of the rising tone. The low tone is

likely to appear as a low fall in contrast with the high fall of the falling tone.

Two types of phonetic context perturb the ideal shapes of the tones. Voiceless initial consonants induce a higher start of the F_0 contour, while voiced initial consonants induce a lower start (Gandour, 1974; Erickson, 1974). This kind of perturbation seems to have little effect on the phonetic integrity of the five tones, although it may serve as a supplementary cue to the voicing state of the initial consonant. It has been argued by historical linguists (Li, 1977), with some perceptual support from recent experiments on Thai (Abramson and Erickson, 1978), that through the phonemicization of these perturbations, the tones of Proto-Tai increased from three to the present-day sets of five or more in the modern languages of the family.

The phonetic context that causes greater deviations from the ideal tonal shapes is that of neighboring tones. In a series of tones spoken without pauses, tonal coarticulation occurs. Although physiological studies of Thai tones (Erickson, 1976) have yet to be extended to sequences, we can infer from acoustic evidence (Abramson, 1979) that this kind of coarticulation is manifested through the overlap of the effects of motor commands for the control of the laryngeal tensions and aerodynamic forces used.

Two sequential effects must be discriminated from tonal coarticulation. First, certain unstressed CV syllables with short /a/ which have low or high tones in citation form are normally toneless in running speech. Another view is that the high and low tones on these syllables are neutralized, and the resulting pitch is assigned to the mid tone. This conclusion is handy for transcription, but the physical evidence suggests instability with F_0 values dominated by the contours of the neighboring lexical tones. The other sequential effect to be excluded from consideration is tonal sandhi. The phonology of the language dictates that when certain kinds of morphemes are conjoined to form compound words, the lexical tone of one of the morphemes is replaced by another tone.

Sentence Intonation and Tones

As one listens to spoken Thai, whether it be an animated conversation or a phlegmatic technical explanation, it becomes clear that in addition to emotional states such linguistic features as

sentence accent and signs of major syntactic breakpoints can be expressed prosodically. The distinction between a statement and a question can also be expressed. In my present approach to the topic, I must lean mainly on my own extensive auditory but limited instrumental observations, as very few useful insights are found in the literature. It would be helpful if native Thai linguists or phoneticians gave more attention to the matter.

As a data base for such observations, as I am ready to make, I have used two kinds of speech material. One is a conversation between two Thai adults of about one minute in length, recorded by J. Marvin Brown for a textbook published by the American University Alumni Association Language Center in Bangkok, Thailand. The other is a monologue recorded by me of the dean of a faculty at a university in Bangkok; speaking for a bit more than a minute, he talks about a new academic program.

Computer-implemented analysis yielded displays of root-mean-square amplitude, wave forms, and F_0 contours. Cepstral analysis was used to extract the fundamental frequency. A sample set is shown in Fig. 1 for the female speaker in the dialogue. Here, by the way, can be seen an example of tonal coarticulation. The phrase /nǎ bǎn/ 'in front of the house' bears two falling tones. The F_0 of the first one does not fall as far as the second; this presumably facilitates the resetting of the larynx for the sharp rise and fall of the second falling tone.

To handle the non-emotive aspects of sentence prosody in Thai, my examination of the present corpus of utterances, reinforced by the arguments of Rudaravanija (1965), leads me to posit three terminal junctures: rising pitch, sustained pitch, and falling pitch. These junctures function at clause ends and sentence ends. They may also function wherever the speaker pauses. The presence of a juncture affects the phonetic shape of the lexical tone on the last one or two syllables. The rising and falling junctures are likely to appear at the end of a breath group. In earlier work (Abramson, 1962) I also posited two pitch registers, high and normal, as units for Thai intonation. I now doubt the relevance of such registers for the non-emotive aspects of sentence prosody in the language. Indeed, to capture emotive prosodic variation, a somewhat more elaborate scheme might be needed. Although, as shown by Noss (1972) and Thongkum (1976),

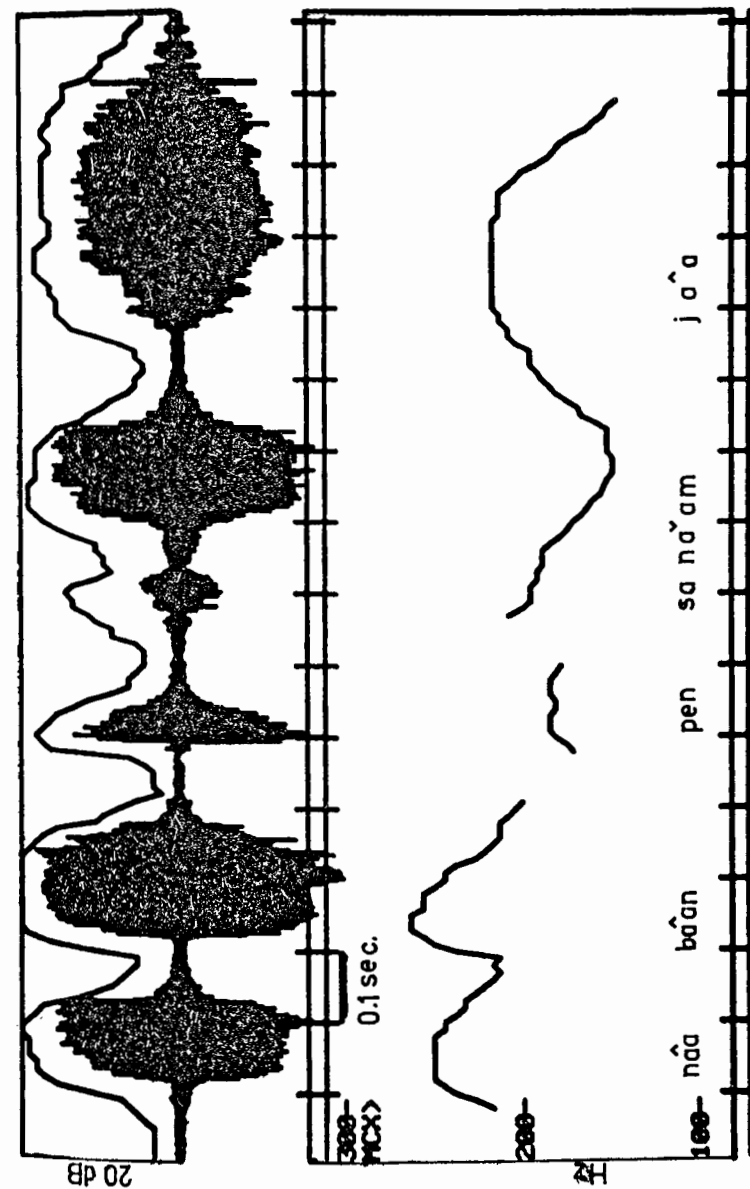


Fig. 1. From top to bottom: R.M.S. amplitude, wave form and fundamental frequency of a Thai woman's production of the sentence /nǎ bǎn pen sa nǎm jǎ/ 'There is a lawn in front of the house.'

rhythmic factors play a role in Thai sentence prosody, they are excluded here because of the scope set by the organizers of the Congress.

Henderson (1949) has argued that aside from the general melodic line of Thai intonation, the "sentence tone" as a whole is mainly determined by the speaker's choice of particles, most of them final particles. She describes seven such sentence tones. Without entering into the question of how many sentence tones there might be, I can at least say that these particles, which indicate, e.g., the sex of the speaker and something about the social relation between the speaker and the hearer, are prime carriers of the terminal junctures. Each particle as a lexical item has a tone of its own in citation form; this tone is usually predictable from the spelling. I doubt, however, that in running speech these "lexical" tones have any standing. The actual pitch imposed on a particle or, sometimes, a sequence of two particles, seems to be determined by the intonation of the whole sentence culminating in a terminal juncture. The resulting "tones" on these particles can sometimes be aligned with the lexical tones of Thai phonology but more often they are deviant; some linguists, apparently in the grip of the view that every Thai syllable must bear a phonemic tone, feel constrained to write each particle with one of the five tones.

In both colloquial and formal discourse, many a sentence contains no particles, so the terminal junctures appear on the final word of the clause or sentence. Fig. 1 shows such an effect. The falling tone on /jɑ̀/ 'grass' at the end of the sentence is considerably lower both at its high point and low point than the two falling tones at the beginning. Even the rising tone just before it on /sanɑ̀m/ 'field' does not rise to a point much higher than the immediately preceding mid tone on /pen/ 'be'. With such a short utterance it is hard to decide whether we have a final falling juncture on the compound word for 'lawn' or a falling intonation contour on the whole sentence.

Sentence accent is manifested by one or more of the following factors: (1) lengthening of the syllable, (2) a tonal contour that approaches the form of the ideal tone, and (3) an increase in amplitude. In the sentence in Fig. 1 the final syllable appears to bear the sentence accent, using factors (1) and

(2). In the phrase /nà bàn/ at the beginning of the sentence, the second syllable is stressed, using factors (1) and (3); the amplitude trace is flattened at the top of the available 20-dB range, indicating saturation.

The points made so far have been descriptions of gross F_0 contours. A problem in intonation analysis is how to present quantitative data that go beyond overall "tunes." The prosodic constructs of the linguist often elude the measuring devices of the phonetician. With the simple-minded analysis for non-emotive prosody into three terminal junctures as a framework, I have made an initial tabulation of frequency movements for such clear examples of terminal juncture as I could find in the corpus. To provide for reasonable comparability of speakers, I treated frequency shifts at terminal junctures as percentages of the voice range. The maximum and minimum F_0 values for each of the three speakers are given in Table 1. Although the speech in both samples was

Table 1

	Voice Range in Hz		
	Dialogue		Monologue
Speakers:	A*	B**	U.W.*
Spread:	130-290	90-235	85-160
Range:	160	145	75
	*Woman	**Man	

calm, the narrower range for the monologue may not be due so much to the habits of that speaker as to the rather dispassionate and thoughtful nature of his discussion compared to the more animated dialogue.

The juncture of sustained pitch is generally found at syntactic breaks where the overall pitch of the voice neither rises nor falls before a brief pause; with or without a pause, the final syllable is prolonged. I have used this sustained pitch as a neutral reference from which to track the movements of the other two junctures. Examining both samples by ear and by eye, I accepted as valid tokens of the three junctures only those instances that were quite unambiguous. This cautious procedure yielded the small number of data in Table 2. The juncture of rising pitch signals surprise, doubt or a question. (Questions can also be marked by means of particles and other morphemes without terminal rising pitch.) The terminal fall appears at the ends of sentences

Table 2
Average Shift Through Voice Range for Terminal Pitch Junctures

Rising		Sustained		Falling	
N	%	N	%	N	%
6	30	14	0*	27	25

*Neutral reference point.

and some major clauses. The "shift" for the sustained pitch is set at 0% as a neutral reference level, while the other two junctures are entered as departures from that neutral level. The data are averaged across the three speakers. None of the tokens of these junctures happened to occur with the low lexical tone.

Even away from the junctures intonation has great effects on the realizations of the tonal phonemes. If the ideal forms of the tones have any psychological validity, then the forms in the sample of running speech have undergone severe distortion. A full account is beyond my reach here. At the same time, as I look at the contours and listen to the speech, I find preservation of the full system of five tones in running speech. That is, the usual linguistic scheme is not an artifact of the formal analysis of the linguist concentrating on citation forms only. Excluded from this generalization, however, must be all particles occurring at major syntactic breaks; they generally have their pitch determined by the sentence intonation without the involvement of lexical tones. Other frequently used function words, such as modals and pronouns, often undergo tonal replacement.

Conclusion

The phonemic tones and sentence prosodies of Thai interact in a rather complicated fashion. Three terminal pitch junctures, often occurring on particles, carry much of the intonation. Although the lexical tones are much influenced in their F_0 movements by sentence intonation, the contrasts between them are preserved except for certain small sets of morphemes. Sentence prosody allows for sentence accent. As in non-tonal languages, it is possible in Thai to use pitch junctures for the difference between statements and at least some kinds of questions.

References

Abramson, A. S. (1962): The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments, Bloomington, Indiana: Indiana U. Res. Center in Anthropology, Folklore and Linguistics.

- Abramson, A. S. (1979): "The coarticulation of tones: an acoustic study of Thai", in Studies in Tai and Mon-Khmer Phonetics in Honour of Eugenie J. A. Henderson, V. Panupong, P. Kullavanijaya and T. Luangthongkam (eds.), Bangkok: Indigenous Langs. of Thailand Res. Proj.
- Abramson, A. S. and D. M. Erickson (1978): "Diachronic tone splits and voicing shifts in Thai: some perceptual data", Haskins Labs. Status Report on Speech Research SR-53, Vol. 2, 85-96.
- Erickson, D. (1974): "Fundamental frequency contours of the tones of Standard Thai", Pasaa 4, 1-25.
- Erickson, D. M. (1976): A Physiological Analysis of the Tones of Thai, Ph.D. diss., University of Connecticut.
- Gandour, J. (1974): "Consonant types and tone in Siamese", JPh 2, 337-350.
- Henderson, E. J. A. (1949): "Prosodies in Siamese: a study in synthesis", Asia Major N.S. 1, 189-215.
- Li, F.-K. (1977): A Handbook of Comparative Tai, Honolulu: U. Press of Hawaii.
- Noss, R. B. (1972): "Rhythm in Thai", in Tai Phonetics and Phonology, J. G. Harris and R. B. Noss (eds.), 33-42, Bangkok: Central Inst. of English Language.
- Pike, K. L. (1948): Tone Languages, Ann Arbor, Mich.: U. of Michigan Press.
- Rudaravanija, Panninee (1965): An Analysis of the Elements in Thai that Correspond to the Basic Intonation Patterns of English, Ed. D. diss., Teachers College, Columbia U.
- Thongkum, T. L. (1976): "Rhythm in Thai from Another View Point", Pasaa 6, 144-158.

WORD PROSODY AND SENTENCE PROSODY IN SWEDISH

Gösta Bruce, Phonetics Department, Lund University, Sweden

Introduction

The present paper summarizes the research on Swedish prosody reported in Bruce (1977) and Bruce and Gårding (1978), and also presents some new ideas concerning intonation in Swedish. The main topic is the relation between word accent, sentence accent and sentence intonation as signalled by F_0 . Our results suggest that observed F_0 -contours typical of statements in four prosodically distinct dialect types (see e.g. Gårding 1975) represent the combined result of one common sentence intonation command, similar word accent commands with different timings, and different sentence accent commands.

Sentence accent

In a prosodic typology for Swedish dialects combining word and sentence prosody (Bruce and Gårding 1978) we have shown how four prosodic dialect types (south, central, east and west) can be characterized mainly by differences in sentence accent. This is illustrated in Figure 1, which shows F_0 -contours of the sentence Man vill lämna våra långa nummer (They want to leave some Långa-numbers), where the placement of sentence accent has been varied. The two accented syllables (accent II) in the sentence - belonging to the verb and the nominal compound respectively - and the secondary-stress syllable in the compound are all surrounded by unstressed syllables. We regard this form of the sentence as optimal for revealing prosodic dialect differences, since the tonal commands can be developed freely with no obvious interference from adjacent commands.

In the south and central dialects, the sentence accent command appears to be superimposed on the word accent command (see Figure 1). Sentence accent is signalled by a wider F_0 -range for the word accent in focus than for the non-focal word accent. For south this wider range is obtained in final position by raising the actual word accent peak, and in non-final position also by lowering the subsequent valley. For central, the wider F_0 -range is achieved mainly by raising the word accent peak in focus both in final and non-final position.

For east and west the sentence accent command comes after

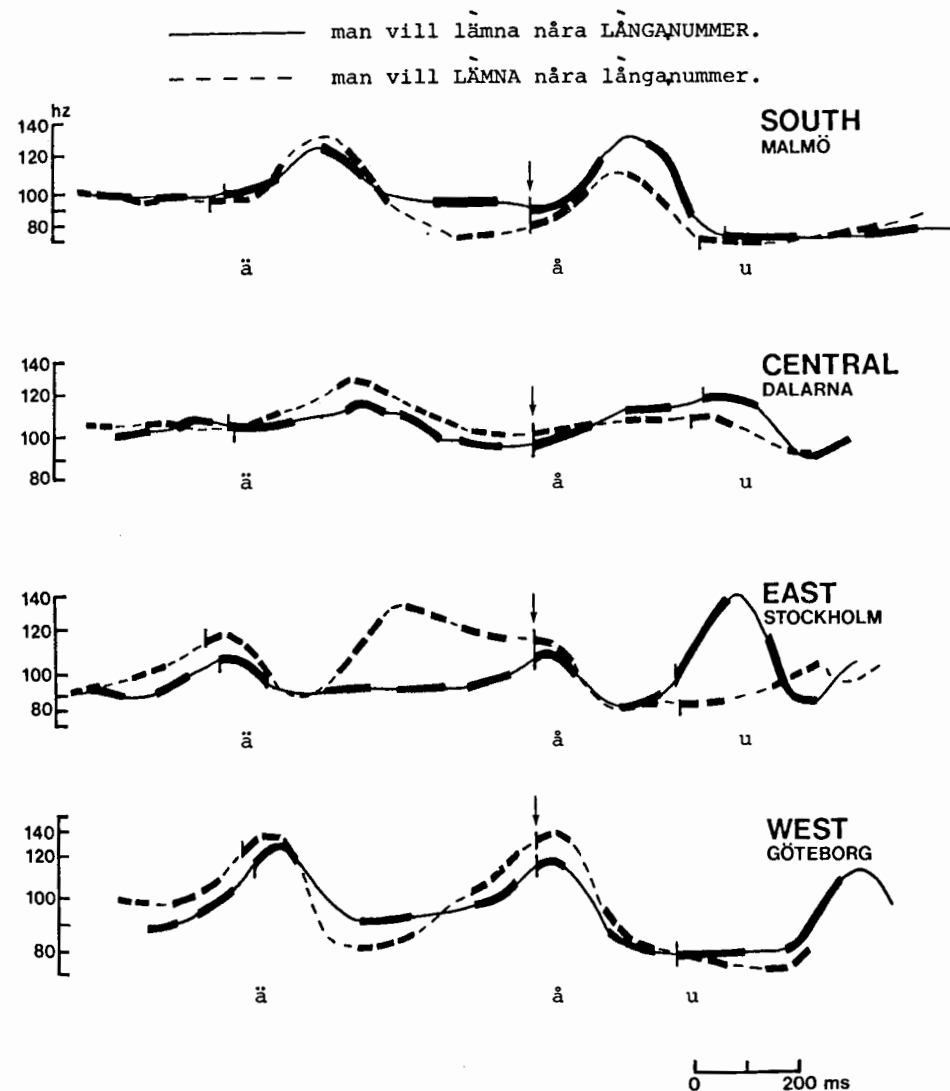


Figure 1. The effect of the placement of sentence accent. F_0 -contours of a sentence with two accented words (accent II). The line-up point (arrow) is at the CV-boundary of the second stressed syllable. Vertical bars indicate the CV-boundaries of the stressed syllables. Vowel segments are drawn in thick lines and consonant segments in thin lines. Focus is indicated by capital letters and non-focus position by small letters.

the word accent command. For east, this means the addition of a separate sentence accent peak immediately after the word accent in focus. For a compound, however, the second peak is postponed until the secondary-stress syllable. The Fo-peak may become a plateau, since in final position Fo will stay on a high level until the utterance-final syllable and in non-final position until the following accented syllable.

For west, the addition of a sentence accent peak in final position occurs in the utterance-final syllable independently of the prosodic structure of the word in focus. In non-final position sentence accent is signalled by both lowering the valley after the word accent in focus and raising the peak of the post-focal word accent, i.e. by a wider Fo-range after the word accent in focus.

This means, in summary, that sentence accent is characterized by a wide Fo-range in all dialect types. In south and central this wide range cooccurs with the word accent in focus, while in east and west it occurs with a time lag.

Word accent

When the contribution of sentence accent to the Fo-contour has been isolated, the word accents appear more clearly. Figure 2 shows Fo-contours of the two word accents (accent I and accent II)

in a non-focal position. It appears that for each dialect type the word accent distinction is signalled mainly by the different timing of the word accent peak relative to the stressed syllable (cf. Haugen 1949). The relative timing difference with accent I always preceding accent II is common to all dialects, but the absolute timing of the word accent peaks varies with dialect. The order of timing between dialects from the earliest to the latest absolute timing of the word accent peaks is east, west, south and central (see Figure 2). For east, at the one extreme, the accent I-peak appears as early as in the pre-stress syllable, while for central, at the other extreme, it occurs in the final part of the stressed syllable. The accent II-peak occurs for east in the initial part of the stressed syllable and for central in the post-stress syllable. The timing of the word accent peaks for west and south falls in between these two extremes.

Sentence intonation

While the dialect-specific features of Swedish intonation

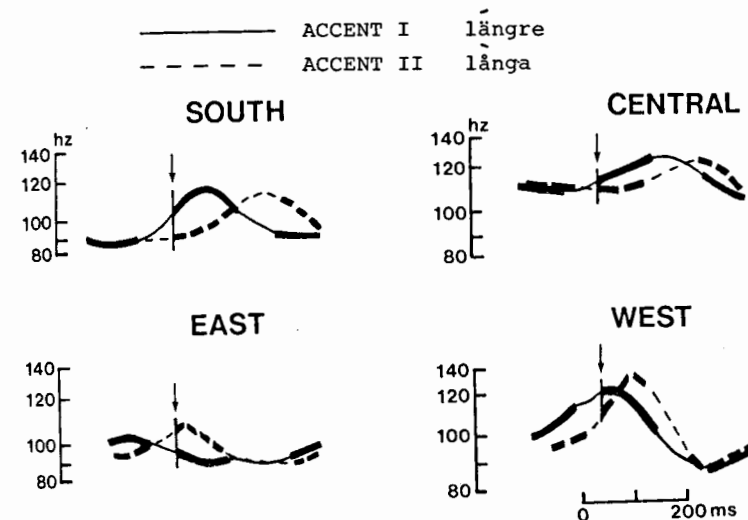


Figure 2. The word accent distinction in non-focal position. Fo-contours of accent I and accent II before focus. The line-up point is at the CV-boundary of the stressed syllable.

are mainly found in the Fo-correlates of word accent and sentence accent, certain aspects of sentence intonation seem to be independent of dialect. Here only statement intonation will be treated. But it appears that also the main features of question intonation are more or less dialect-independent (see Gårding forthcoming).

A characteristic feature of sentence intonation in many languages of the world is the downdrift of Fo over the course of an utterance, also referred to as the declination effect (see e.g. Cohen and 't Hart 1967). This means, in general, that Fo is higher in the beginning than at the end of an utterance with each Fo-peak and Fo valley being lower than the preceding one. In Swedish the topline connecting successive peaks of an utterance appears to decrease at a faster rate than the baseline connecting successive valleys, which means that the Fo-range is also gradually decreasing. This Fo-downdrift is found in English and Danish, too (see Breckenridge 1978, Thorsen 1978).

A model was proposed in Bruce and Gårding (1978) to account for the main features of sentence intonation in Swedish. In my experience it is not typical of the Fo-downdrift in Swedish to be

evenly distributed over an utterance. The total Fo-drop in an utterance for a given speaker appears to be the same, however, regardless of the length of the utterance. The actual course of the Fo-downdrift in an utterance seems to be dependent on several factors, such as the location of sentence accent and of the word accents. Figure 3 illustrates this point. It shows Fo-contours of the sentence *Man vill lämna våra långa nunnor* (They want to leave some tall nuns) containing three accented syllables (accent II) with varying placement of sentence-accent.

The Fo-drop appears to have a stepwise and not gradually decreasing course. The downstep takes place in connection with the accented syllable. In unaccented syllables before and after a word accent there is no systematic downward slope.

It will be assumed that the basic sentence intonation command (statement intonation) has a stepwise decreasing course with a successive narrowing of the range. Sentence accent normally interferes with sentence intonation, introducing a break into the basic pattern (see Figures 1 and 3). This may affect the course of the topline as well as that of the baseline depending on the dialect. Before focus the Fo-drop and the narrowing of the Fo-range of the word accents appear to be relatively gentle. After focus, however, it usually decreases more rapidly, with a considerable narrowing of the Fo-range.

As a consequence of the Fo-downdrift there is a position-dependent variation of word accent and sentence accent. A word-accent in the beginning of an utterance has higher Fo-values for peak, valley, and range than at the end (everything else being equal). Also the corresponding sentence accent values tend to decrease with position in the sentence.

Conclusion

Sentence intonation (statement intonation) in Swedish can be represented independently of dialect by a stepwise decrease of Fo taking place in connection with the word accents and affecting peak, valley, and range values. This downdrift pattern will often be locally disturbed by sentence accent introducing a break into the basic Fo-course in a dialect-specific way. This suggests that the downdrift is linguistically controlled rather than a consequence of some peripheral production constraint. Instead it can be assumed to be built into the intonation system.

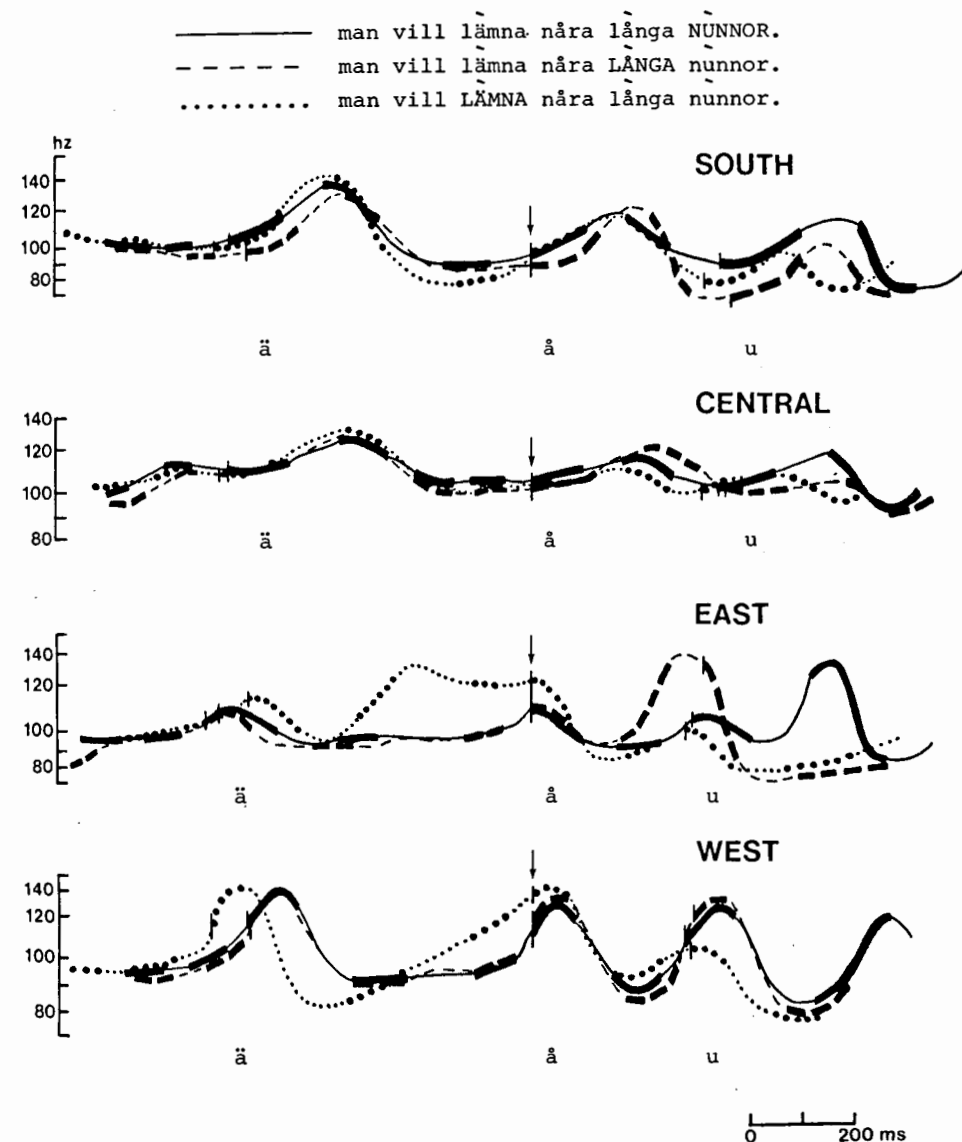


Figure 3. Downdrift in Swedish - the combined effect of statement intonation and the location of sentence accent and word accent. Fo-contours of a sentence with three accented words (accent II). The line-up point is at the CV-boundary of the second stressed syllable.

Acknowledgements

This work was carried out within the project "Swedish prosody" in cooperation with Eva Gårding. It was sponsored by the Swedish Humanistic and Social Sciences Research Council.

References

- Breckenridge, J. (1978): Declination as a phonological process, Department of Linguistics and Philosophy, MIT, Cambridge, Massachusetts, Mimeographed.
- Bruce, G. (1977): Swedish word accents in sentence perspective, Lund: Gleerup.
- Bruce, G. and E. Gårding (1978): "A prosodic typology for Swedish dialects", to appear in Nordic prosody, E. Gårding (ed.), Lund: Gleerup.
- Cohen, A. and J. 't Hart (1967): "On the anatomy of intonation", Lingua 19, 177-192.
- Gårding, E. (1975): "Towards a prosodic typology for Swedish dialects", The Nordic languages and modern linguistics 2, 466-474, K.-H. Dahlstedt (ed.), Stockholm: Almqvist & Wiksell.
- Gårding, E. (1977): The Scandinavian word accents, Lund: Gleerup.
- Gårding, E. (forthcoming): Sentence intonation in an accent language.
- Haugen, E. (1949): "Phoneme or prosodeme?", Language 25, 278-282.
- Thorsen, N. (1978): "Aspects of Danish intonation", to appear in Nordic prosody, E. Gårding (ed.), Lund: Gleerup.

ON THE INTERACTION OF ACCENTUATION AND INTONATION IN DUTCH

J. 't Hart and R. Collier, Institute for Perception Research,
P. Box 513, Den Dolech 2, Eindhoven 5612 AZ, The Netherlands

Introduction

The general aim of this symposium is to track down universal features concerning the relation between word prosody and sentence prosody, with the exclusion of durational phenomena. We will present our viewpoints about the relation at issue in as far as we have been confronted with it in our experiences with Dutch intonation and accentuation.

Dutch intonation lacks the occurrence of "tonemes". Therefore, on the level of the word, we can limit our discussion to phenomena of "lexical accentuation"; on the level of the sentence we have to discuss "sentence accents" and "intonation". On the latter level particular problems may arise as regards the interaction of accentuation and intonation. In fact, sentence accents manifest themselves as "pitch accents" on words or syllables, whereas intonation patterns are realized as "pitch contours" extending over entire utterances. In other words, two aspects of sentence prosody are interwoven in the same phonetic variable, viz. the variation of F_0 (or pitch) as a function of time.

In the literature, the problem of the interaction between accentuation and intonation is coped with in a number of ways, most of which share the assumption that the overall pitch contour can be considered simply as the sum, or the linear addition, of the variations of F_0 associated with accentuation and those associated with intonation. We will confront this assumption with a number of phenomena as observed in Dutch.

Word prosody

In Dutch, each word is said to possess a lexical accent. In polysyllabic words the location of this accent is not fixed but, in principle, it can be predicted by rule.

Lexical accents may be considered as abstract features on the level of word phonology. A subsequent and separate problem is to specify when and how these lexical accents manifest themselves phonetically.

If we take, for example, the word Amerika (America), we find a lexical accent on the second syllable. When this word is spoken in isolation, as a one-word utterance, a listener will hear the

second syllable as prominent. Acoustic measurement of the utterance will reveal substantial changes of F_0 (hereafter referred to as "pitch movements") on the second syllable: a rise, a fall, or a combination of the two. These pitch movements may be considered a phonetic correlate of the lexical accent, since it can be shown experimentally that their deletion or displacement causes prominence judgments to change accordingly. That pitch movements are efficient cues for prominence is not surprising if one realizes that, psycho-acoustically speaking, only a few percent change of F_0 is sufficient to be supraliminal, whereas in actual speech F_0 changes are observed that are eight to ten times as large as these threshold values.

At this point we may conclude that lexical accents in one-word utterances are realized (among other things) by means of pitch movements: rises, falls, or combinations of both, apparently without a particular preference for any of these various possibilities.

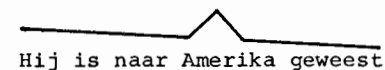
Since a one-word utterance is an utterance all the same, we may as well extend the discussion to longer utterances with one or more accents.

Sentence prosody

In the introduction we have already mentioned that on the level of the sentence, pitch changes correlate with both accentuation and intonation. On the linguistic level these categories are easily kept apart, but on the phonetic level the distinction may become blurred to the extent that the observable pitch changes can be associated with either of the two categories (or with both). One would therefore like to sort out how accentuation and intonation interact in shaping up the ultimately observable course of the pitch in concrete utterances.

Let us illustrate this problem with the following example. The Dutch sentence Hij is naar Amerika geweest (He has been to America) may be pronounced with one accent, viz. on the syllable -me- of Amerika. Again, F_0 measurements will show pitch movements on that syllable, e.g. a rise-fall combination. Of course, a sufficiently refined F_0 measurement will also reveal changes on other syllables than the accented one, but experiments with artificial, stylized pitch contours show that such changes are not

relevant to the perception of either accentuation or intonation. In the example below, the stylized rise-fall may be preceded and followed by a gradual downward running movement of pitch (the so-called "declination"). This is sufficient to make the contour prosodically well-formed.



The essential shape of the pitch contour of this example corroborates our prior suggestion that, phonetically, no distinction can (nor has to be) made between the realization of a lexical accent in a one-word utterance and the realization of a single accent in a longer utterance: both may manifest themselves in the form of the same pitch movement(s). In both cases the pitch movements are the phonetic correlate of the accent inasmuch as their deletion or displacement has consequences for the perception of prominence.

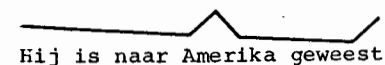
Primacy of accentuation

Let us now consider one possible explicit formulation of the principle that the overall pitch contour is obtained by a linear addition of the accentual and intonational requirements. The principle would then be phrased as follows:

P1 Those pitch movements that co-occur with prominent syllables are entirely and exclusively related to accentuation, the remaining pitch movements of the contour are associated with intonation.

How would P1 account for the pitch contour of Hij is naar Amerika geweest? P1 might relate either the rise or the fall to accentuation and the other pitch movement (plus the declination) to intonation; or P1 would assign the rise-fall combination to accentuation, in which case there would be no particular intonation (except for the declination).

The sentence Hij is naar Amerika geweest may also be pronounced as follows:



In this instance, P1 would assign the final rise to intonation,

Variant:

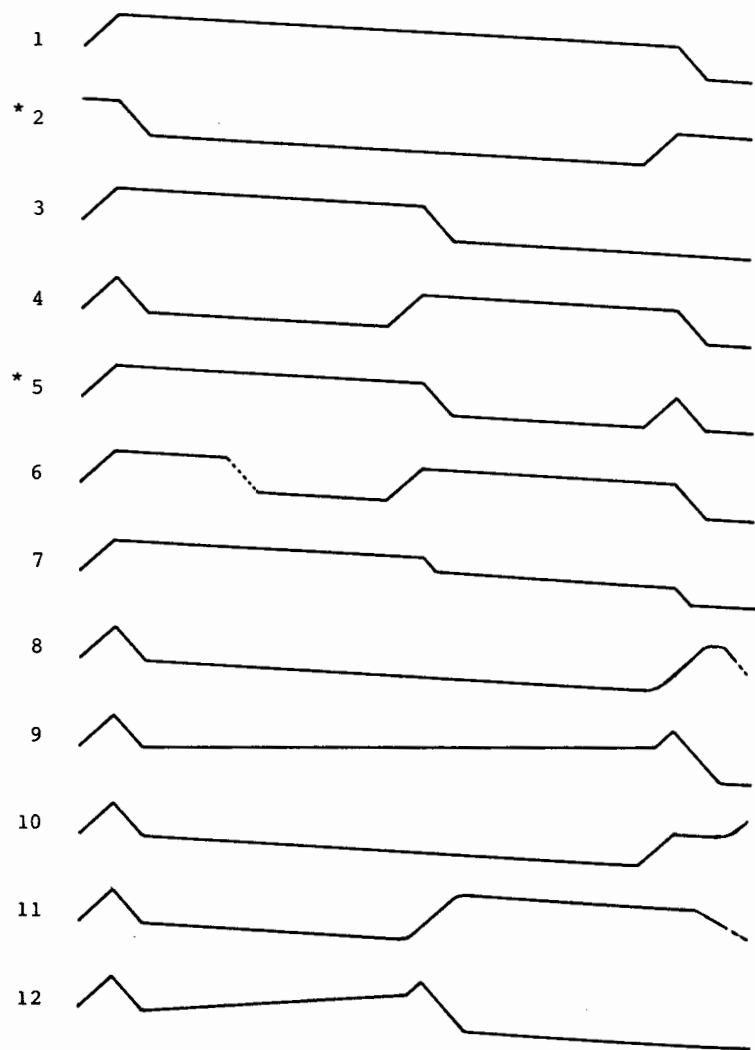


Fig. 1. Stylised representation of twelve different pitch contours for a given specimen sentence. Variants 2 and 5 are not acceptable.

since it does not lead to an additional accent.

Let us now turn to another, more complicated Dutch specimen sentence, with variants of accentuation and intonation. These variants are listed in Figure 1. They all refer to possible ways of intoning the sentence Grootmoeder gaat met de kinderen naar het zwembad (Grandmother goes with the children to the swimming-pool).

In variant 1 the syllables groot- and zwem- are to be accented. This can be brought about by having the pitch go up on the syllable groot- and down on the syllable zwem-.

Since, apparently, it is possible to produce accent by a mere rise or fall, we might examine the case of variant 2, with a fall on the syllable groot- and a rise on the syllable zwem-. Such a contour can easily be constructed and made audible by means of a speech synthesizer or an Intonator. It appears that the contour of variant 2 sounds unacceptable to Dutch ears. This means that the choice of the kind of pitch movement that has to take care of accentuation is not free.

P1, then, is unsatisfactory since it does not account for the choice of the accent lending pitch movements. There is nothing in the nature of the accents themselves that would predict this choice. So, our suggestion is that the choice of the kind of accent-lending pitch movements is subordinate to the kind of intonation pattern that is to be realized.

An alternative: primacy to intonation

This suggestion of subordination is not reflected in the formulation of P1. On the contrary, P1 assumes a primacy in favour of the pitch movements needed for accentuation. A logical alternative would take the primacy of intonation as a starting-point. Basing ourselves on this primacy, we will first formulate an alternative principle, P2, and then try to present empirical support for it.

- P2
- The nature and the order of all the pitch movements in an utterance are determined by the intonation pattern.
 - Among the pitch movements of any intonation pattern there is at least one which possesses such phonetic properties as are necessary for bringing about a pitch accent.
 - The location of the accent-lending pitch movement(s) is

determined by the position of the words that carry sentence stress, and more specifically, by the position of the lexically accented syllable in each of these words.

This formulation does not allow any movement to be entirely and exclusively related to accentuation. Therefore, the addition principle is abandoned.

Returning to variant 1, we would, on the basis of P2, interpret its pitch contour as follows: the intonation pattern which is realized consists essentially of an accent-lending rise and an accent-lending fall, in that order; their locations are in accordance with the accentual demands.

We will now check whether such an account would also be applicable to other accentual and (or) intonational variants.

P2 would predict that if the accentuation requirements change, the only change(s) will be in the location of the accent-lending pitch movements. Suppose, for instance, that the same intonation pattern as in variant 1 is used, but that instead of the syllable zwem-, the syllable kin- has to be accented. P2 will predict a rise and a fall in that order, the rise on groot- and the fall on kin-. This gives rise to variant 3, which is indeed a possible, and well-formed contour.

A special case is provided in sentences with only one accent. If still the same intonation pattern is to be used, then the rise and the fall must necessarily coincide on the one accented syllable. This accounts for the example Hij is naar Amerika geweest.

Yet another illustration of the same intonation pattern with different accentuation is shown in variant 4, where three accents are at stake. In such a case, the introduction of one additional accent-lending pitch movement would in principle be sufficient. However, since the essential property of the intonation pattern being used is "a rise followed by a fall", there is no other possibility than a repetition of these two movements. These may be combined on one of the accented syllables, but not just on any of them. Indeed, an intonational requirement in Dutch is that the separated rise and fall should only occur on the penultimate and the last accented syllable, respectively. This requirement is violated in variant 5, which is therefore unacceptable.

The rise and fall that were introduced in variant 4 to account

for the additional accent need not coincide on the same syllable. The accentuation requirement can also be met by means of a mere rise on the syllable groot-. The fall may then occur somewhat later, as in variant 6. In such a case it coincides with a word boundary (and more specifically with a major syntactic boundary). The fall at the word boundary cannot give rise to an additional pitch accent, due to its particular location, but it may serve another purpose, viz. the marking of a syntactic boundary.

All this goes to say that the intonational requirements are met in such a flexible way that the accentual (and sometimes also syntactic) demands are satisfied at the same time. P2 accounts for this flexibility.

We have shown before that P1 cannot account for the unacceptability of variant 2. P1 would be capable of accounting for the structure of variants 3, 4, 5 and 6. But it cannot explain why variant 5 is unacceptable, nor why variants 4 and 6 are melodically dissimilar, while remaining accentually identical.

If P1 is unsatisfactory to a limited extent in the case of a number of accentual variations in contours based on one intonation pattern, its complete failure becomes apparent whenever variations of the intonation pattern are at stake. This can be illustrated by means of variants 7 to 10, in which four different intonation patterns are used, again with accents on the syllables groot- en zwem- (and in variant 7 also on kin-). P2 accounts for the variants 7 to 10 by stating that when different intonation patterns are realized, some of their essential components fall into such places as is necessary to accommodate the pitch accents. P1, however, could never explain how the accent on e.g. the syllable zwem- is phonetically manifested in so many mutually exclusive ways. Again, if the location of one of the accents changes, the accent-lending pitch movement that figures among the indispensable components of the intonation pattern is shifted to a different position, as appears from the comparison of variant 11 to 8, and of 12 to 9.

Conclusion

The examples given show that pitch accents can have quite a number of different forms of appearance. If, according to P1, one would assume that the pitch movements associated with accentuation

are entirely autonomous and do not, in any respect, constitute a part of the essential components of the intonation pattern, this leads to the following difficulty: how will one predict for every single accent which type(s) of pitch movement should be used to realize it? The ability to make such predictions is a real necessity since the various kinds of pitch movement cannot follow one another in an arbitrary order in an utterance.

Therefore, the most satisfying way to account for the combinatory restrictions among pitch movements associated with accents is to assume that they are determined by the chosen intonation pattern. This is expressed by P2, in which, contrary to P1, accentuation is subordinate to intonation.

In our opinion, P2 also accounts for the interaction between pitch accents and intonation patterns in other "accent languages" without tonemes, such as English, German, and others.

SOME OBSERVATIONS ON THE PERCEPTION OF STRESS IN CZECH

Přemysl Janota, Institute of Phonetics, Charles University,
Prague, Czechoslovakia

Considerable attention has been paid to the problems of the perception of stress during the last three decades. Generally speaking, the results of the various experiments have largely coincided in showing that there is no one-to-one correlation between judgments of stress and any single physical feature of the speech signal.

In an earlier experiment we tried to determine the influence of three physical dimensions - intensity, fundamental frequency, and duration - on the perception of stress in Czech; the sound material consisted of synthetic disyllabic items. In listening tests of this kind, the listeners' judgments are based, generally speaking, on two complexes of phenomena, which may be labelled as, firstly, acoustic properties of the speech signal and, secondly, contextual cues. The relation between the two complex factors can vary within considerable limits in natural utterances; in experimental conditions it is possible, however, to suppress, to a certain extent, the part played by one of the factors. We also treated the selection of a suitable test word as important. In the experiments we used throughout combinations of the syllable 'se': the word 'sese', meaning 'session', does exist in Czech, though it is quite rare - this means, it is less likely to evoke the association of a word with first syllable stressed, and even less likely in the light of the fact that an actual sequence of syllables se-se is very common in Czech.

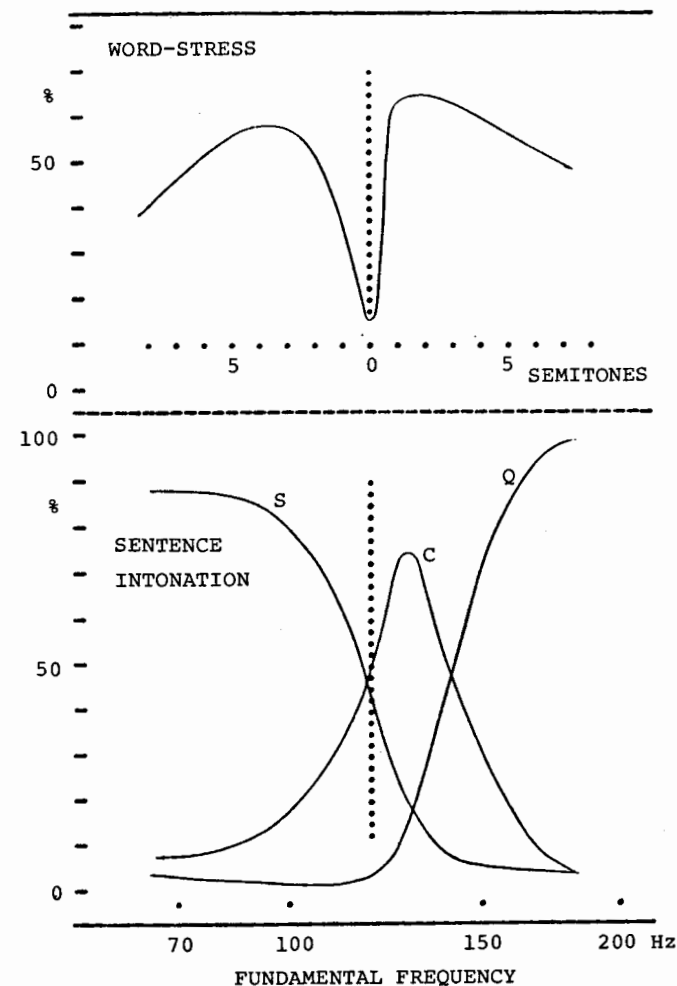
In the instructions to the test, the listeners were invited to mark stressed syllables; in English terminology the term 'prominence' would perhaps be more fitting, but the corresponding term does not exist in Czech phonetic literature. In the last edition of the standard handbook of Czech orthoepy (Hála 1967, 65) it is stated: 'By stress we usually understand the phonetic emphasis of one syllable with respect to others; within a single word this emphasis is called word-stress.' The task of the listeners was then to mark this 'phonetic emphasis', prominence. The test stimuli were produced by means of a simple synthesizer of our own construction, recorded on tape in random order and presented to 170 Czech

listeners. The results of the test can be summarized as follows: increasing intensity of the stimulus leads to an increase in the number of 'stressed syllable' judgments; increased duration has a similar effect. With changes of the fundamental frequency the relationship was found to be somewhat different. Relatively small changes - a semitone up or down - led to a conspicuous increase in the number of judgments 'stressed syllable', while a further raising or lowering of the fundamental frequency did not lead to any further increase in the number of 'stressed' judgments, but on the contrary led to a decrease.

In the evaluation of the results of this experiment we were aware that they were valid for the experimental conditions of the test and for the synthetic material used. However, we have left as an open question the unexpected influence of frequency changes on stress evaluations, i.e. the stronger effect of small changes of the fundamental frequency and the similar effect of both the increase and decrease of the fundamental frequency on the number of judgments 'stressed'. One hypothesis here was that more marked changes of tonal pitch are evaluated rather as sentence melody. Besides, a possible influence of the synthesizer had to be taken into consideration.

In the following tests we concentrated specially on these problems. Firstly, a test was prepared in collaboration with J. Liljencrants at the Speech Transmission Laboratory of the Royal Institute of Technology in Stockholm and this time, the OVE III synthesizer was used. 42 test sentences 'byla to sese' - meaning 'it was/was it a session' - were prefabricated, in which the fundamental frequency values of the last vowel were systematically changed with much closer graduations than in the earlier test. The changes in the fundamental frequency corresponded to 0, $\frac{1}{2}$, 1, 2, 4, 6, and 8 semitones in each direction away from the fundamental frequency of the previous syllable; in addition, the intensity of the last vowel was changed at 3 levels. This material was then prepared for a new listening test with other groups of listeners: firstly, the isolated stimuli 'sese' were extracted from the recording and again the listeners' task was to mark the syllables they thought stressed; in the second test, the whole synthetic sentences were used and the listeners' task was to indicate with each item, whether they felt the sentence to be

statement, question or continuative sentences. A total of 100 listeners took part in the test, and the results are presented in the following graph:



Upper part of the graph: percentages of judgments 'stress on the second syllable'

Lower part of the graph: percentages of judgments S-statement, Q-question, C-continuative sentence

Scale in semitones: difference in pitch of the second syllable of the test word

It can be seen from the graph that the evaluation of stress accords well with the previous experiment: a steep rise from a dip of the curve in the middle with a slow decrease in percentages 'stressed' farther away from the center line. The graph also very clearly shows that in the area of fundamental frequency changes where the number of judgments 'stressed' begins to fall, there is a marked distinction of sentence melody - in the lower part of the graph - when it comes to evaluating the 'sentence' stimulus. By and large, these results corroborate the hypothesis that smaller changes of the fundamental frequency contribute more to the perception of stress, while greater intervals are more at play in the domain of sentence intonation.

In this test again the listeners responded readily and within narrow limits to signals of identical fundamental frequency; these stimuli with no or only a very small difference between the first and the second syllable were, however, not only identified, but the judgment 'first syllable stressed' was ascribed to them quite consistently. A possible interpretation of this finding is that the listeners evaluated both the rise and fall in frequency as a deviation from a pitch level of the first syllable in the test word, held constant throughout the whole test. This would be in agreement with findings from analyses of running speech, where departures from the basic contour of intonation in the direction up or down both are used in Czech to express prominence of a part of an utterance.

Two different tests were then prepared to investigate this hypothesis: in one test an attempt was made to keep the characteristics of the test similar to those of the previous experiment with the exception of the fundamental frequency of the first syllable of the test word: this was changed in a random order within an octave interval. The test items were prepared by means of the synthesizer HO 2, constructed by Maláč et al. (1975), and the listening tests were finished and evaluated in August 1978. In another test again the sequence 'sese' was used, but this time as a natural speech signal. In view of the high occurrence rate of 'se', even in iteration, in Czech texts it was easy to prepare a continuous text, a short story, which contained within its two pages a total of 116 repetitions of the syllable 'se' with the stress assumed to fall variously on the first or on the second

syllable of the sequence se-se. A professional speaker read the story and the tape recording was then used to prepare test tapes containing copies of all the sese combinations cut out of the master tape. A total of 50 listeners then heard the isolated items; their task was as with the earlier tests.

The results of the perception tests with both synthetic and natural items in isolation compare well with those of the earlier tests with synthetic stimuli, with a very clear difference, however, in the perception of changes of pitch: in the present experiments the effect of a change, i.e. the influence of pitch rise on an increase of judgments 'stressed', is manifest primarily towards the upper pitch levels. Clearly, this difference is not due to any difference inherent in the use of synthetic or natural speech signals, but to the different way in which the experiment was organized. In the first two experiments the fundamental frequency level of the first syllable of the test word was constant and hence probably provided a reference level comparable with a basic contour of sentence intonation. - In short, the results of the experiment with the isolated items of natural speech can be described as follows: the syllables marked 'stressed' had a higher fundamental frequency than those marked 'unstressed' in 93% of the cases, a higher peak intensity in 79%, and a longer duration in 44% of the cases.

Conclusions

By means of listening tests using disyllabic items, the influence of changes in intensity, fundamental frequency, and duration on the perception of stress in Czech can be shown.

In general, it can be demonstrated that an increase of any of these parameters leads to an increase of the number of judgments 'stressed'.

With changes of the fundamental frequency, however, the growth of the judgments 'stressed' is distinctly non-linear: slight changes of approximately a semitone lead to a considerable increase in the number of judgments 'stressed', while larger changes have a lesser effect on the evaluation of stress in a group of Czech listeners.

Results of a test with identical synthetic stimuli once in isolation and then in a simple sentence context corroborate the hypothesis that small changes of fundamental frequency have a

stronger effect on word-stress evaluations, whereas more marked changes have a noticeable effect on evaluations of sentence intonation.

The influence of context may be strong even in tests with isolated items: in tests with a constant pitch of the first syllable of disyllabic test items, deviations - both up and down - from this constant level are found to increase the number of syllables marked as stressed. In tests without the constant pitch level of the first syllable, only increase in fundamental frequency is found to add to the number of judgments 'stressed'.

Listening tests with isolated items only have been referred to in the present paper; it is obvious that even here there is a strong tendency of the listeners to evaluate the items as parts of a broader context.

Bibliography

- Bolinger, D.L. (1958): "On intensity as a qualitative improvement of pitch accent", *Lingua* 7, 175-182.
- Fry, D.B. (1958): "Experiments in the perception of stress", *L&S* 1, 126-152.
- Fyodorova, N. (1973): "The effect of some acoustic parameters of the synthetic speech signal on the perception of stress by Russian listeners", in *Speech Analysis and Synthesis III.*, W. Jassem (ed.), 229-248, Warsaw: PWN.
- Gårding, E. and A.S. Abramson (1965): "A study of the perception of some American English intonation contours", *SL* 19, 61-79.
- Hadding-Koch, K. and M. Studdert-Kennedy (1964): "An experimental study of some intonation contours", *Phonetica* 11, 175-178.
- Hála, B. (1967): *Výslovnost spisovné češtiny I.*, Praha: Academia.
- Janota, P. (1967): "An experiment concerning the perception of stress by Czech listeners", *Acta Universitatis Carolinae - Philologica* 6, *Phonetica Pragensia*, 45-68.
- Janota, P. and J. Ondráčková (1975): "Some experiments on the perception of prosodic features in Czech", in *Auditory Analysis and Perception of Speech*, G. Fant and M.A.A. Tatham (eds.), 485-496, London: Academic Press.
- Janota, P. and Z. Palková (1974): "The auditory evaluation of stress under the influence of context", *Acta Universitatis Carolinae - Philologica, Phonetica Pragensia IV.*, 29-59.
- Jassem, W., J. Morton and M. Steffen-Batóg (1968): "The perception of stress in synthetic speech-like stimuli by Polish listeners", in *Speech Analysis and Synthesis I.*, W. Jassem (ed.), 289-308, Warsaw: PWN.
- Lehiste, I. (1970): *Suprasegmentals*, Cambridge, Mass.: MIT Press.
- Lieberman, P. (1960): "Some acoustic correlates of word stress in American English", *JASA* 32, 451-454.
- Maláč, V., M. Ptáček, P. Dvořák and B. Borovičková (1975): "The HO 2 system - a digitally controlled terminal analog synthesizer", *Tesla Electronics*, 121-124.
- Mol, H. (1972): "The investigation of intonation", *Acta Universitatis Carolinae - Philologica* 1, *Phonetica Pragensia III.*, 176-178.
- Morton, J. and W. Jassem (1965): "Acoustic correlates of stress", *L&S* 8, 159-181.
- Ondráčková, J. (1961): "On the problem of the function of stress in Czech", *Zs.f.Ph.* 14, 45-54.
- Rigault, A. (1970): "L'accent dans deux langues à accent fixe: le français et le tchèque", in *Prosodic Feature Analysis*, P.R. Léon, G. Faure, A. Rigault (eds.), 1-12, Montréal: Didier.
- Rigault, A. and T. Arkwright (1972): "Les paramètres acoustiques de l'accent en tchèque", *Proc.Phon.* 7, 1004-1011.
- Romportl, M. (1973): "On the synonymy and homonymy of means of intonation", in *Studies in Phonetics*, M. Romportl, 137-146, Prague: Academia.
- Uldall, E.T. (1962): "Ambiguity: question or statement? or 'Are you asking me or telling me?'"', *Proc.Phon.* 4, 779-783.

WORD STRESS AND SENTENCE STRESS IN VARIOUS TONE LANGUAGES

Eunice V. Pike, Summer Institute of Linguistics
7500 West Camp Wisdom Road, Dallas, Texas 75236 USA

Introduction

The nine languages summarized here all use two or more tones as part of the features which contrast lexical items. All but one of the languages, Fasu (4), show tonal contrasts on both stressed and nonstressed syllables. In Fasu the tonal contrasts occur on stressed syllables only. One language, Mikasuki (6), contrasts long versus short vowels in addition to tone.

Instead of contrasting lexical tone in the prepause syllable, Tenango Otomi (3) has a different pitch system on that syllable, one that indicates the attitude of the speaker.

Diuxi Mixtec (9) has two types of word stress, one marked by vowel length, and the other by allotones. Some words have both types, but others have only the type marked by vowel length. Eastern Popoloca (2) also has two types of word stress, one marked by vowel length and the other by consonant length. In this language the two types never occur in the same word.

Probably in all of the languages, loudness is optionally present on stressed syllables. Vowel length is used to mark stress in at least five of the languages. Consonant length is one of the features marking stress in at least three languages.

By the term "word stress", I mean the syllable which is the nucleus of a rhythm wave, in this case, the phonological word (Pike 1976, 54-69). By "sentence stress" I mean the nucleus of a larger rhythm wave. As I have used it in this paper, this rhythm wave coincides with the pause group (K. Pike 1967, 392-403), so by "sentence stress" I mean the syllable which is the nucleus of a pause group.

A word uttered in isolation is between pauses; therefore the stressed syllable of a word in isolation has sentence stress. To identify word stress, usually there must be at least two words in the utterance. Word stress is the stress which remains on a word when it occurs in the margin of a pause group.

If, in a specific language, word stress and sentence stress occur on the same syllable, it is perhaps impossible to know which features are marking word stress when the words are studied only in isolation. It may be, for example, that in Mikasuki (6) some of the features which were described as those of word stress were actually features of sentence stress, since the data, for the most part, were studied from words uttered in isolation.

There is less apt to be confusion when sentence stress occurs on a different syllable from that where word stress occurs. For example, in Ayutla (7) and Acatlan Mixtec (8), and also in Tenango Otomi (3) word stress occurs on the first syllable of the stem, but sentence stress (with some exceptions) occurs on the prepause syllable.

Perhaps the thing that surprised me most, as I summarized the nine languages, was that in only one of the languages, Fasu (4), were vowel allophones spoken of as being determined by their occurrence in relation to a stressed syllable. If I have the opportunity to hear these languages again, that is one of the points I will check.

Languages summarized

(1) Marinahua of Peru (Pike and Scott 1962) has a contrastive tone, high versus low, on each syllable (p.7). Word stress occurs on the first syllable of the stem and is marked by vowel length (p.4). Sentence stress occurs on the first syllable of the stem

of the last word in the sentence (p.2); it has an even longer vowel than occurs with word stress. Sentence stress is also marked by allotones, that is, the high is raised, and the low tone usually glides down when in a syllable with sentence stress (p.8). When the speaker is irritated, the last consonant of the sentence may be lengthened, and the loudness may shift from the normally stressed syllable to the prepause syllable (p.4).

(2) In Eastern Popoloca of Mexico (Kalstrom and Pike 1968), contrastive tone consists of four level tones plus ten tone clusters which are combinations of those tones. There are two types of stress. Some words have stress marked by a long vowel (that stress occurs on the next to the last syllable of the stem, p.16,18). Other words have a stress that is marked by a long consonant (that stress usually occurs on the last syllable of the stem (p.17). When only one consonant occurs in the syllable, that consonant is lengthened. When a consonant cluster occurs, it is the /h/, /?/, or /n/ of the cluster which is lengthened--all consonant clusters have either /h/, /?/, or /n/. All words have either one stress type or the other, but no words have both.

In a normal sentence, the prepause syllable is usually short, louder than other syllables in the sentence, and it ends in a sharp glottal stop (p.28). A polite sentence is raised in key and the prepause syllable is long, lenis, gradually getting softer as it glides upward. There is no final glottal stop in that type of sentence (p.29).

(3) In Tenango Otomi of Mexico (Blight and Pike 1976), high, low and upgliding tone contrast lexical items. The upgliding tone occurs only on syllables with word stress. Word stress occurs on the first syllable of the stem, and is marked by vowel length

(p.56); it is also marked in that voiceless stops are preaspirated when they are initial in a stressed syllable (p.52). A low tone in a stressed syllable is slightly lower than a nonstressed low; a stressed high usually has a slight downgliding allotone when occurring between voiced consonants (p.55). Sentence stress occurs on a prepause syllable and is marked by loudness. It is that prepause syllable that carries contrastive intonation. There is no contrast of lexical tone on the prepause syllable nor on any word-final syllable (p.55).

(4) In Fasu of Papua New Guinea (May and Loeweke 1965), a contrast of high versus low tone occurs on only one syllable per word--the stressed syllable; the placement of that syllable is not predictable. Vowels /i/ and /e/ have open variants which fluctuate with close variants when prestressed if the stressed vowel is /i/ or /e/ respectively (p.92). Within a sentence, there is a gradual downdrift of pitch. In a question-doubt sentence, without an interrogative marker, there is a small upglide on the final syllable (p.95). Other attitudes of the speaker may be indicated by a wider spread in the tone levels, or by voice quality, etc.

(5) In Golin of Papua New Guinea (Bunn 1970), there is contrastive tone, high versus low, on each syllable. Word stress occurs on the final high, or if there is no high, then on the final low (p.4). Sentence stress occurs on the same syllable as word stress, but it is louder and if it is a syllable with high tone, it usually has a higher allotone. When sentence stress occurs on a syllable with low tone, there is optionally a lower allotone (p.5). Sentence stress may occur on any word of the sentence; it is usually used for emphasis (p.6).

(6) Mikasuki of Florida USA (West 1962) has contrastive long versus short vowels. Mikasuki also has three levels of tone and one tone cluster which are used in lexical items (p.82). In a question, one of the syllables of the sentence may have an extra-high tone and additional length, and one of the words may end in glottal stop (p.82,89). (Glottal stop is not lexically pertinent.) Other clause types, negative statements, imperatives, for example, may also have the extra-high tone or tone clusters that do not occur in simple lexical items (p.89). Word stress usually occurs on the highest nonfinal syllable (p.85). In the normal sentence, there is a gradual drop of pitch between words (p.88-89).

(7) In Ayutla Mixtec of Mexico (Pankratz and Pike 1967), there are three levels of tone, contrastive on each syllable (p.291). Word stress occurs in the first syllable of the stem and is marked by loudness and also by allophones of the consonants. That is, when contiguously following the stressed syllable, voiceless stops and affricates are preaspirated, voiced continuants are lengthened, voiceless continuants are either lengthened or preceded by a slight hiatus (p.288). Allotones mark word stress in that a proclitic with high tone is not as high as a stressed high which immediately follows it (p.291). Throughout the sentence there is a downdrift of pitch. Sentence stress is usually louder than other syllables, and occurs either on the prepause syllable, or on the syllable with word stress-- there is variation in accordance with the CV pattern and the tone sequence of the last word (p.294).

(8) Acatlan Mixtec (Pike and Wistrand 1974) has contrastive tone on each syllable: low, mid, high and up-step (p.83). Word

stress occurs on the first syllable of the stem. It is marked by allophones of consonants (p.100,103) in that all consonants, except the flapped /r/, are lengthened when contiguously following a stressed syllable. If the syllable which follows stress does not begin with a consonant, then it is the stressed vowel which is lengthened. Sentence stress occurs on the syllable immediately preceding pause (p.104). There is general downdrift in relaxed speech in that each low tone tends to be lower than the preceding low (p.84).

(9) Diuxi Mixtec (Pike and Oram 1976) has two contrastive tones, high versus low, one of which occurs on each syllable. There are two types of word stress. The type marked by a lengthened vowel occurs, on each word, on the first syllable of the stem (p.322). The second type of word stress occurs on the word-final syllable, but only on some words. It is marked by allotones. That is, a stressed high between another high and pause has a sharp downglide. When between low and pause, the high may not downglide, but it is definitely higher than a nonstressed high in that environment. A stressed low tone downglides from a point starting noticeably higher than a nonstressed low (p.325).

References

- Blight, Richard C. and Eunice V. Pike (1976): "The Phonology of Tenango Otomi", IJAL 42, 51-57.
- Bunn, Gordon and Ruth (1970): "Golin Phonology", (Papers in New Guinea Linguistics XI) Pacific Linguistics A 23, 1-7.
- Kalstrom, Marjorie and Eunice V. Pike (1968): "Stress in the Phonological System of Eastern Popoloca", Phonetica 18, 16-30.
- May, Jean and Eunice Loeweke (1965): "The Phonological Hierarchy in Fasú", Anthropological Linguistics 7.5, 89-97.
- Pankratz, Leo and Eunice V. Pike (1967): "Phonology and Morphotonemics of Ayutla Mixtec", IJAL 33, 287-99.
- Pike, Eunice V. (1976): "Phonology", in Tagmemics I, Aspects of the Field, Ruth M. Brend and Kenneth L. Pike (eds.), 45-83.
- Pike, Eunice V. and Joy Oram (1976): "Stress and Tone in the Phonology of Diuxi Mixtec", Phonetica 33, 321-33.
- Pike, Eunice V. and Eugene Scott (1962): "The Phonological Hierarchy of Marinahua", Phonetica 8, 1-8.
- Pike, Eunice V. and Kent Wistrand (1974): "Step-up Terrace Tone in Acatlán Mixtec (Mexico)", in Advances in Tagmemics, Ruth M. Brend (ed.), 81-104, Amsterdam: North Holland Pub. Co.
- Pike, Kenneth L. (1967): Language in Relation to a Unified Theory of the Structure of Human Behavior, (2nd revised ed.) The Hague: Mouton.
- West, John David (1962): "The Phonology of Mikasuki", Studies in Linguistics 16, 77-91.

LEXICAL STRESS, EMPHASIS FOR CONTRAST, AND SENTENCE INTONATION
IN ADVANCED STANDARD COPENHAGEN DANISH

Nina Thorsen, Institute of Phonetics, University of Copenhagen

Due to lack of space, no references will be made to the very considerable literature on intonation in other languages, nor will any extensive documentation be given.

1. A model of Danish intonation

Intonation in short sentences in Advanced Standard Copenhagen (ASC) Danish may be presented as in fig. 1, which is but a model, with the advantages and shortcomings that modeling almost always entails in terms of simplicity and inaccuracy, respectively. It is based on recordings by six subjects, three males and three females, of a rather elaborate material (Thorsen 1978a, 1979b). The qualitative statements which can be read off the figure are perfectly representative of all the subjects, but the quantifications involved are, of course, averages, and no one subject behaves as mathematically neatly as the model would have you believe.

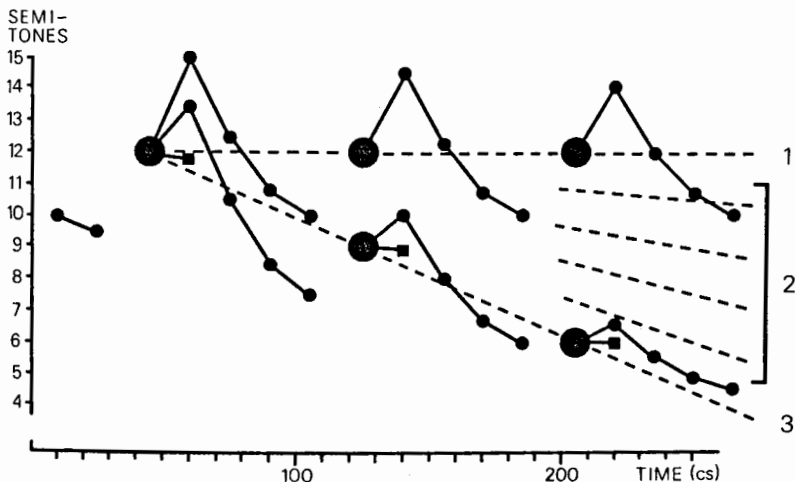


Figure 1

A model for the course of F_0 in short sentences in ASC Danish. 1: statement questions, 2: interrogative sentences with word order inversion and/or interrogative particle, and non-final periods (variable), 3: declarative sentences. The large dots represent stressed syllables, the small dots unstressed ones, and the small squares represent an unstressed syllable being the only one between two stressed ones (see further the text). The full lines represent the F_0 pattern associated with stress groups, and the broken lines denote the intonation contours.

A basic assumption underlying fig. 1 is that the complex course of fundamental frequency (Fo) in an utterance is the outcome of a superposition of several components: (1) A sentence component which supplies the INTONATION CONTOUR (broken lines). (2) On the contour is superposed a stress group component which furnishes the STRESS GROUP PATTERNS (full lines). (3) To the resultant of those two components is added, in words containing stød, a stød component, rendering STØD MOVEMENTS. However, as stød words had been excluded from the material, the model does not include this particular feature. These first three components are language specific and thus "speaker controlled". (4) Finally, intrinsic Fo level differences between segments, and coarticulatory variations at segment boundaries supply a MICROPROSODIC COMPONENT, which is not consciously controlled by the speaker, but due to inherent properties of the speech production apparatus and which, therefore, is superfluous in the model from the point of view of the human speaker. - This concept of "layers" in intonation is anything but original; the triviality of the statement does not, however, deprive it of its validity or its relevance: firstly, it is tremendously useful in the interpretation of Fo tracings (Thorsen forthcoming) and, secondly, it has a very direct bearing on the theme of this symposium, 'The relation between sentence prosody and word prosody (stress and tone)':

The relation between word stress and sentence prosody (i.e. sentence intonation: duration and intensity are not considered) is physically a very close-knit and intricate one, but on the higher and more abstract levels we may hypothesize that very little of the mutual influence which is customary in a relationship takes place, as long as we are dealing with neutral lexical stresses.

1.1 Stress group patterns

As can be seen from the full lines of fig. 1, the stress group pattern can be described as a (relatively) low stressed syllable followed by a high-falling tail of unstressed syllables. This pattern is a predictable and recurrent entity, though allowing for contextual variation in the magnitude of the rise from stressed to unstressed syllable and in the slope of the fall through the unstressed ones. It was this observation which gave rise to the definition of the STRESS GROUP in ASC Danish as A STRESSED SYLLABLE PLUS ALL SUCCEEDING UNSTRESSED SYLLABLES (within the same, non-compound sentence), irrespective of intervening word or morpheme boundaries, and, as a consequence of this predictability and recurrency, it al-

so brought about the definition of the INTONATION CONTOUR as THE COURSE DESCRIBED BY THE STRESSED SYLLABLES ALONE.

1.2 Intonation contours

The intonation contours tend to vary systematically with sentence type, declarative sentences having the most steeply falling contours, at one extreme, and statement questions (i.e. questions with statement syntax where only the intonation contour signals their interrogative function) having "flat" contours, at the other extreme. In between these two are found other types of questions as well as non-final periods. For a further account of these contours and their perception, see Thorsen 1978b, 1979a.

2. Implications of the model

2.1 Fo movements in syllables

The model does not specify the Fo movements of syllables: the tonal composition of the stress group pattern as one of LOW plus HIGH FALLING allows for a very simple account of Fo movements in vowels and consonants: segments do not carry specific movements (except when stød is involved) but simply float on the Fo pattern, and slight variations in Fo movement would be due then to the fact that segments do not always hit the patterns at exactly the same place.

2.2 The course of the intonation contour

(a) When the number of stress groups changes, everything else being equal, so does the slope of a given contour, leaving only the "flat" ones intact; the constancy presumably lies in the interval between the first and the last stressed syllable, with intervening stressed syllables evenly distributed between them, and not in a certain rate of change (this point needs further verification which I hope to present orally in August).

(b) When the number of unstressed syllables varies in the stress groups, the stressed syllables will not be equidistantly spaced in time, and the straight lines of fig. 1 break up into a succession of shorter ones with unequal slopes.

Combining the effects of changes of both types leaves us with an infinity of physically different intonation contour configurations. On a higher level in production these variations may not exist, and perceptually they may be obliterated, turning the contours into smoothly slanting slopes, (1) if what we aim at producing and what we perceive are equal intervals between stressed syllables and not the actual slope of the contour, and (2) if we assume that isochrony, be it not a physical reality, is a psycho-

logical reality with the speaker/listener.

2.3 Fo patterns of stress groups

2.3.1 Stress groups with more than one unstressed syllable

(a) In statements, the rise from stressed to unstressed syllable is, on the average $1\frac{1}{2}$, 1, and $\frac{1}{2}$ semitone, respectively, in the first, second, and third stress group. In statement questions, the rises amount to 3, $2\frac{1}{2}$, and 2 semitones, respectively. The difference in magnitude of this rise, between patterns riding on different contours is very likely a direct consequence of differences in the level of the following stressed syllable.

(b) The decrease with time in the rise from stressed to unstressed syllable is the same in statement questions and statements, one semitone. This decrease, which is independent of the particular contour, may be seen as a consequence of either of two distinct processes, or of a combination of them. It may be a "voluntary" decrease, i.e. a signal of finality, and/or it may be a physiological phenomenon: the closer you get to the end of the utterance, the less "energy" is expended and the less complete the gestures will be; either or both phenomena may also account for the less and less steep falls through the unstressed syllables.

If the variation in the Fo patterns with intonation contour and time is physiologically determined, the speaker may be unconscious of it, and the listener may neglect or compensate for it.

2.3.2 Stress groups with only one unstressed syllable

Stress groups with one unstressed syllable will of course be shorter than those with several, a feature which is not reflected in fig. 1. - A single unstressed syllable does not accomplish a full rise-fall when the following stressed syllable is considerably lower than the preceding one, as is the case in statements. Instead it lands on very nearly the same level or slightly below the preceding stressed syllable and, accordingly, the rise-fall is amputated. A full rise-fall may be intended by the speaker and the amputation be due to a shortcoming in the peripheral speech production mechanism. Accordingly, the listener may well re-introduce a rise-fall (this is, indeed, my own subjective impression). But we have here an indication that time (rhythm) overrides Fo when the two are in conflict. On the other hand, there is definitely a tendency towards as complete rise-falls as possible. Two unstressed syllables will traverse more than half the fall exhibited by four, everything else being equal. - These two facts together are

another reminder that speech is not a card-board structure but a smooth and dynamic process.

2.4 Conclusion

If the assumptions made about production and perception of Fo courses hold water, we are left with two components which physically are highly interactive but on more abstract levels may be invariant, apart from the fact that contours change with sentence type.

3. Emphasis for contrast

By emphasis for contrast is meant the extra prominence on one of the syllables in the utterance, used to denote a contrast which may be implicit or may be explicitly stated in the context. I have deliberately avoided terms like 'focus', 'sentence accent', or 'nucleus' because these terms are used, in a number of languages, to describe a phenomenon different from emphasis for contrast: one of the lexically stressed syllables in the utterance will always have slightly greater prominence (realized, very roughly speaking, as a more elaborate Fo movement within that syllable), and if nothing else is specified by the context, it will fall on the last stressed syllable. - A similar phenomenon does not exist in ASC Danish as a thing apart from emphasis for contrast. Whenever and wherever such a slightly heavier stress is introduced, it invariably invokes the impression of contrast. Insofar as we are not faced with incomplete evidence or with a false dichotomy, i.e. one due to differences in concepts, Danish seems to be markedly different from e.g. English, German, and Swedish.

3.1 Contrastive stress and Fo

The following account is based on a material of sentences, uttered in dialogues, where the contrasts were all explicitly stated in the context (but I strongly believe that they would have looked no different had they been implicit). - When emphasis for contrast occurs, it affects the intonation contour as well as Fo patterns.

Fig. 2 compares the neutral edition with three statements where the emphasis lay on the first, second, and third lexically stressed syllable. (Durational differences between the neutral and emphatic editions are very slight and there is no doubt that Fo is the prime cue to contrast, as it is to neutral lexical stress.) The obvious changes introduced in the Fo course by emphasis for contrast is a raising of the syllable in question (represented by a star), a drastic fall from first to second unstressed syllable, plus a not inconsiderable shrinking of the surrounding Fo patterns:

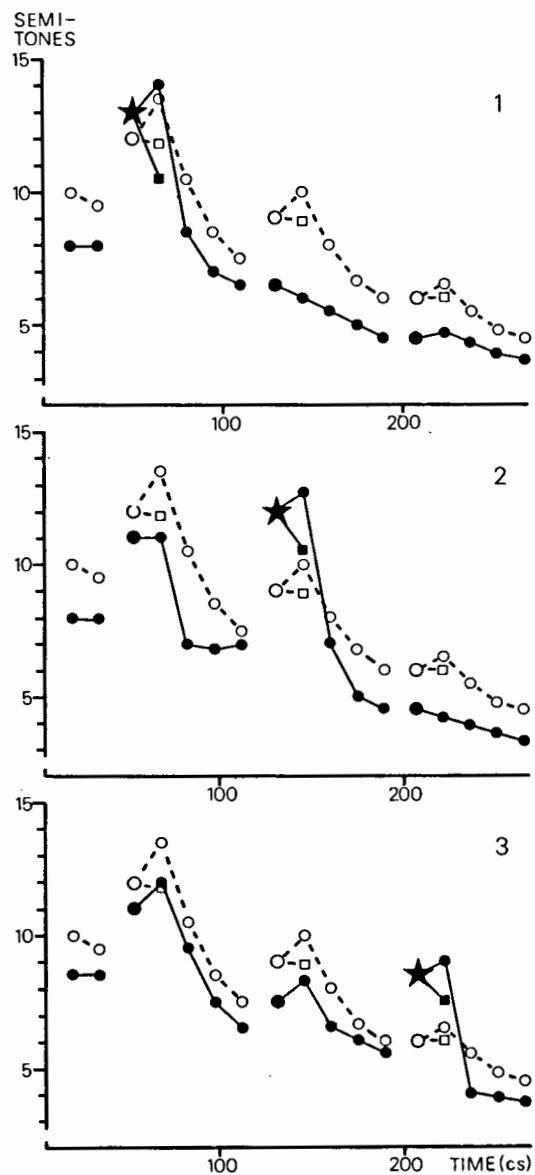


Figure 2

Models for statements with emphasis for contrast on (1) the first, (2) the second, and (3) the last stressed syllable, compared to the neutral edition (broken lines and empty dots).

(1) when emphasis is on the first stressed syllable, it is higher, the rise to the first unstressed syllable is smaller, and the fall through the following unstressed syllables is steeper than for the neutral case. The levels of the second and third stressed syllables are considerably lowered and the LOW+HIGH FALLING pattern is annihilated in the second and shrunk in the third stress group. The syllables of the second and third stress groups look, tonally, more like a series of unstressed syllables continuing the fall in the first one.

(2): emphasis on the second stressed syllable repeats the pattern of (1), and we also get a certain reduction of the first stress group with a very steep fall from first to second unstressed syllable.

(3) the pattern repeats itself in the last stress group, with a shrinking of the preceding ones as well.

Again we note that a single unstressed syllable does not accomplish a full rise-fall but instead drops well below the preceding emphatic one.

The feature common to

the three cases seems to be that the syllable on which the emphasis for contrast occurs must stand out clearly from the surroundings, which is brought about by a raising of that syllable as well as by a lowering of the immediate surroundings, except for the first of several post-tonic syllables. The change is slightly greater in the succeeding than in the preceding Fo course. During some informal experiments performed with the ILS-system for analysis and synthesis at the Institute of Linguistics, Uppsala University, it appeared that shrinking the Fo course in the surroundings is sufficient to create the impression of emphasis for contrast. To get emphasis on the word 'sidste' in the statement 'Det er sidste bus til Tiflis.' (It's the last bus for Tiflis.) it is sufficient to change the rise from 'bus' to 'til' to a level or a slight fall, whereas just raising the stressed syllable of 'sidste' will not do the trick. Likewise, to get emphasis on 'bus', lowering the unstressed syllable of 'sidste' will do and just raising 'bus' does not accomplish anything.

The three Fo courses in fig. 2 look widely different and only vaguely resemble fig. 1 "3" although the utterances still sound declarative. What constitutes the intonation contour in utterances with emphasis for contrast, I hesitate to say at present. They may resemble one-word utterances in that the difference between statement and question lies in the level of and movement within the emphatically stressed syllable as well as in the course of the succeeding unstressed ones (Thorsen, 1978a), or the intonation contour may be extrapolated from, and thus still be definable in terms of, the lexically stressed syllables surrounding the emphatic one. The first solution would be interesting, because it implies that in utterances with emphasis, word prosody takes precedence over sentence prosody, whereas the second solution would make the definition of intonation contour apply to a wider range of utterances.

- References (ARIPUC = Ann. Rep. Inst. Phonetics, Univ. Copenhagen)
- Thorsen, N. (1978a): "An acoustical investigation of Danish intonation", *JPh* 6, 151-175.
 - (1978b): "On the identification of selected Danish intonation contours", *ARIPUC* 12, 17-73.
 - (1979a): "Aspects of Danish intonation", *Travaux de l'Institut de Linguistique de Lund* 13 (in print).
 - (1979b): "Lexical stress, emphasis for contrast, and sentence intonation", *ARIPUC* 13 (in preparation).
 - (forthcoming): "Interpreting raw fundamental frequency tracings of Danish", *Phonetica*.

SENTENCE TONE IN SOME SOUTHERN NIGERIAN LANGUAGES

Kay Williamson, School of Humanities, University of Port Harcourt, Port Harcourt, Nigeria

The lexical tones of words can be modified in various ways:

1. By essentially phonetic rules, such as tone-spreading (Hyman & Schuh, 1974): e.g. Yoruba /Low-High/ → [Low-Rising] because Low 'spreads' into the following High. Such phonetic rules result in phonological change if the conditioning factor is lost.
2. By morphophonemic rules, i.e. rules whose phonetic motivation is no longer obvious: e.g. in the Kolokuma dialect of Izon two words which have the same tone pattern in isolation may have different tonal effects upon the following word (cf. Williamson, 1965).
3. By the interaction of purely tonal morphemes with the tones of the normal morphemes which consist of both segments and tones: e.g. the subject concord marker of Edo is analysed by Amayo (1975) as having lost its segmental features in practically all contexts, so that its presence is normally detected only by its tonal effects on neighbouring morphemes. Purely tonal morphemes appear to be restricted to common grammatical elements.

Lexical tones are also modified to show sentence type. In some languages such modifications involve changing the absolute but not the relative pitch of sentences: e.g. in Kana (Ogoni group: tonemes High, Mid, Low):

Statement: Lo tɔ. [--] 'The house.'
 Question: Lo tɔ? [^^] 'The house?'
 Exclamation: Lo tɔ! [^^] 'The house!'

The basic Mid-Mid tone (seen in the statement) is raised for a question and raised even more for an exclamation (this is indicated phonetically by writing it above the square brackets, i.e. outside the normal voice range). This type of modification is here called intonational, and is regarded as comparable to what obtains in a non-tone language.

Other languages have a second type of modification co-existing with the intonational type. This involves a change of the tone pattern, not simply a general modification of the absolute pitch, and is here called sentence tone. In the examples that follow, the tone system of each language will be summarized and then the sentence tone modifications will be stated.

A. YEKHEE (=ETSAKO), Ekpheli dialect (North-Central Edo group), Elimelech (1976):

Basic tones: H, L

Tone rules: a) downdrift on each series of highs separated by low
 b) falling and rising glides formed from HL, LH
 c) downstep from simplification of rising glide

Sentence tone: (for nouns; verbal sentence questioning said to be different but not specified, Elimelech, 1976, 50):

1. statement: additional final low added to final high
2. question: additional final high added to statement tone pattern

Lexical tone patterns of disyllabic nouns: LL, HL, HH

Data:	Statement	Question
1. LL 'cup'	Àkpà. [--]	Àkpà? [^]
2. HL 'house'	Ówà. [^-]	Ówà? [^-]
3. HH 'axe'	Údzé. [^^]	Údzé? [^^]

B. DEGEMA (Delta Edo group), personal investigation, analysis tentative:

Basic tones: H, L; downstep probably predictable

Tone rules: a) downdrift on each series of highs separated by low
 b) falling glide formed from HL
 c) the final low in a series of lows becomes high (under certain conditions)
 d) all but the first of a final series of highs are downstepped

Sentence tone: 1. statement: basic tones + tone rules
 2. question: final low added to statement tone pattern, combines variously with preceding tone
 3. exclamation: general raising of statement tone

Lexical tone patterns of disyllabic nouns: LL, HH, HL (loanwords only):

Data	Statement	Question	Exclamation
1. LL 'head'	Útóm. [^-]	Útóm? [^-]	Útóm! [^-]
2. HH 'river'	Édá. [^^]	Édá? [^^]	Édá! [^^]
3. HL 'cat'	Pòsì. [^-]	Pòsì? [^-]	Pòsì! [^-]
4. LL (?)	Mòyá. [^-]	Mòyá? [^-]	Mòyá! [^-]
	'He is coming.'	'Is he coming?'	'He is coming!'
5. HH (?)	Àbò. [^^]	Àbò? [^^]	Àbò! [^^]
	'He is there.'	'Is he there?'	'He is there!'

C. ISOKO (Southwestern Edoïd group), Elugbe (1977):

Basic tones: H, LTone rules: a) falling glide formed from HL
b) no downdriftSentence tone: 1. statement: final series of lows raised to mid
2. question: additional final low added
3. exclamation: no raising of final series of lowsLexical tone patterns of disyllabic nouns: LL, HH, HL

Data:	Statement	Question
1. LL 'native doctor'	Òbù. [--]	Òbù? [..]
2. HH 'warrior'	Ógbá. [^^]	Ógbá? [^^]
3. HL 'maize'	Ókà. [-^]	Ókà? [-^]

D. IZON, Kolukuma dialect (Ijọ group), personal investigation:

Basic tones: H, LTone rules: a) downdrift on each series of highs separated by low
b) complex morphophonemic rulesSentence tone: 1. statement: basic tones + tone rules
2. question: slight raising of highs, cancellation of downdrift, final low added
3. exclamation: general raising of highs and of final low; cancellation of downdrift
4. command: slight raising of highs, cancellation of downdriftLexical tone patterns of disyllabic nouns: LH (3 types), HH (2 types), HL

Data:	Statement	Question	Exclamation
1. LH 'yam'	Bùrú. [-^]	Bùrú? [-^]	Bùrú! [-^]
2. HH 'medicine'	Dírí. [^^]	Dírí? [^^]	Dírí! [^^]
3. HL 'sail'	Bálà. [-^]	Bálà? [-^]	Bálà! [-^]
Statement	Question	Exclamation	Command
4. Wónì múdọ̀.	Wónì múdọ̀?	Wónì múdọ̀!	Wómìnì mú!
[- - -]	[- - ^]	[- - ^]	[- - ^]
'We have gone.'	'Have we gone?'	'We've gone!'	'Let's go!'

E. NEMBE (Ijọ group), personal investigation:

Basic tones: H, LTone rules: a) downdrift on each successive high even without intervening low
b) complex morphophonemic rulesSentence tone: 1. statement: final low tone becomes high
2. question: final high tone becomes low
3. exclamation: general raising of highs; cancellation of downdrift after low
4. command: additional final low added to statement patternLexical tone patterns of disyllabic nouns: LH, LL, HL

Data:	Statement	Question	Exclamation
1. LH 'yam'	Bùrú. [-^]	Bùrú? [..]	Bùrú! [-^]
2. LL 'book'	Dírí. [-^]	Dírí? [..]	Dírí! [-^]
3. HL 'Ebi'	Ébí. [-^]	Ébí? [-^]	Ébí! [-^]
Statement	Question	Exclamation	Command
4. Ébí hó.	Ébí hò?	Ébí hó!	Ébí, hóó!
[- - -]	[- - ^]	[- - ^]	[- - ^]
'Ebi came.'	'Did Ebi come?'	'Ebi came!'	'Ebi, come!'

F. KALABARI, didlect of Eastern Ijọ, Jenewari (1977) and personal investigation:

Basic tones: H, L, distinctive downstep (')Tone rules: a) downdrift on each series of highs separated by low
b) complex morphophonemic rulesSentence tone: 1. statement: a) basic tones + tone rules (for non-emphasized nouns)
b) basic tones + (H)'H (first H only after L)

(for verb forms ending H, NPs ending in pronoun/article, and emphasized nouns, especially in answer to a question)

2. question: basic tones + tone rules
3. exclamation: as for 1b), plus general raising
4. command: basic tones + tone rules + additional LLexical tone patterns of disyllabic nouns: LL, HH, H'H, HL, LH

Data:	Statement a) + Question	Statement b)
1. LL 'yam'	Bùrú. Bùrú? [..]	Bùrú'ú. [-^]
2. HH 'book'	Dírí. Dírí? [^^]	Dírí'i. [^^]
3. H'H 'house'	Wá'rí. Wá'rí? [^^]	Wá'rí'i. [^^]
4. HL 'leopard'	Sírí. Sí'rí? [-^]	Sírí'i. [-^]
5. LH 'Gogo'	Gógó. Gógó? [-^]	Gógó'o. [-^]

	<u>Statement b)</u>	<u>Question</u>	<u>Exclamation</u>	<u>Command</u>
6.	ò ɓòtẹ́'ẹ̀	ò ɓòtẹ́?	ò ɓòtẹ́'ẹ̀!	ò ɓòo!
	[_ _ \]	[_ - -]	[_ - \]	[_ \]
	'He has come.'	'Has he come?'	'He has come!'	'Let him come!'

G. IGBO, Green and Igwe (1963) and personal investigation:

Basic tones: H, L, distinctive downstep (')

Tone rules: a) downdrift on each series of highs separated by low
b) falling and rising glides formed from HL, LH
c) morphophonemic rules

Sentence tone: 1. statement: basic tones + tone rules
2. question: a) intonational raising of high, etc., in nominal sentences
b) inseparable subject pronouns change from H to L, in verbal sentences
3. exclamation: a) intonational raising of high and lowering of low
b) cancellation of downdrift
4. command: basic tones + tone rules

Lexical tone patterns of disyllabic nouns: HH, LH, HL, LL, H'H

<u>Data:</u>	<u>Statement</u>	<u>Question</u>	<u>Exclamation</u>
1. HH 'head'	Ísí. [--]	Ísí? [--]	Ísí! [--]
2. LH 'rat'	Òké. [- -]	Òké? [- -]	Òké! [- -]
3. HL 'house'	Úlọ. [- _]	Úlọ? [- _]	Úlọ! [- _]
4. LL 'earth'	Àlà. [- -]	Àlà? [- -]	Àlà! [- -]
5. H'H 'tooth'	É'zé. [- -]	Ézé? [- -]	É'zé! [- -]
6.	Ó jèrè ahíá. [- _ _] [- -]	Ó jèrè ahíá? [- _ _] [- -]	Ó jèrè ahíá! [- _ _] [- -]
	'He went to market.'	'Did he go to market?'	'He went to market!'

Summary and conclusions

The statement is the most basic form of sentence, for:
a) Most commonly the statement form uses only basic forms plus general tone rules (Kana, Degema, Iẓon, Kalabari (type a), Igbo), whereas questions and exclamations usually require extra rules.
b) Questions are normally formed by addition to the statement form (Yekhee), or to the basic/statement form (Degema, Isoko, Iẓon).
c) The only case where the statement form appears more complex than the question is in Kalabari b), and this is apparently an emphatic

form which has become generalized for certain grammatical categories.

d) Exclamations are formed by modification of the statement, never of the question.

CONCLUSION 1. The statement is the most basic sentence type.

This is probably a universal.

Questions are formed from statements by:

- a) addition of a floating low tone (Degema, Isoko, Iẓon)
- b) replacement of a high by a low tone (Nembe, Igbo in verbal sentences)
- c) addition of a floating high tone (Yekhee)
- d) general raising (Kana, Iẓon, Igbo in nominal sentences)
- e) cancellation of downdrift (Iẓon)

CONCLUSION 2. Yes/no questions are marked by a question marker, by intonational raising (including cancellation of downdrift), or by both. This is probably a universal.

Question markers are morphemes which are segmental + tonal or purely tonal (cf. the introduction). The floating low tones of Degema, Isoko, and Iẓon result from morphemes which have lost their segmental features; e.g. in Engenni (Delta Edoid group, closely related to Degema) the sentence-final question marker is à (Thomas, 1969).

It is also possible for the floating tone to replace the adjacent segmental tone (Nembe, Igbo). Historically, the floating tone first combines with the adjacent tone to form a glide (Yekhee, Degema, Isoko, Iẓon); later, the glide is simplified to a level tone:

Nembe: *H + ' > *F > L
Igbo: *' + H > *R > L

CONCLUSION 3. Question markers are morphemes which in a tone language always include tone and sometimes lose their segmental features, after which they are realized like other floating tones. There is no universal that question markers must have high tone, or that they must be sentence-final.

Exclamations are marked by the raising of high tones (sometimes also of low tones) to a greater extent than in questions, and sometimes by the cancellation of downdrift (Iẓon, Nembe, Igbo) and the lowering of low tones (Igbo).

CONCLUSION 4. Exclamations are marked by the raising of tones, especially high ones, and by the increasing of the intervals between tones. This is probably a universal.

Commands seem not to be primarily marked by tone/intonation changes. In the reported cases they either have the same pattern as statements (Igbo) or the statement pattern with an additional floating tone marker (Nembe, Kalabari), or with slight raising and cancellation of downdrift (Izon).

CONCLUSION 5. Commands either resemble statements or differ from them only by the addition of an imperative marker, or by slight raising and elimination of downdrift. This is probably a universal.

References

- Amayo, A. (1975): "The structure of verbal constructions in Edo (Bini)", J. of West Afr. Lang. 10:1, 5-27.
- Elimelech, B. (1976): A Tonal Grammar of Etsako, UCLA Working Papers in Phonetics, 35.
- Elugbe, B.O. (1977): "Some implications of low tone raising in Southwestern Edo", Stud. in Afr. Ling., Supplement 7, 53-62.
- Green, M.M. and G.E. Igwe (1963): A Descriptive Grammar of Igbo, Berlin: Akademie-Verlag and London: Oxford University Press.
- Hyman, L.M. and R.G. Schuh (1974): "Universals of tone rules: evidence from West Africa", Linguistic Inquiry 5, 81-115.
- Jenewari, C.E.W. (1977): Studies in Kalabari Syntax, Ph.D. thesis, University of Ibadan.
- Thomas, E. (1969): A Grammatical Description of the Engenni Language, Ph.D. thesis, University of London.
- Williamson, K. (1965): A Grammar of the Kolokuma Dialect of Ijo, London: Cambridge University Press.

PERCEPTION OF SPEECH VERSUS NON-SPEECH

Summary of Moderator's Introduction

David B. Pisoni, Speech Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA. 02139, U.S.A.

Historically, the study of speech perception may be said to differ in a number of ways from the study of other aspects of auditory perception. First, the signals typically used to study the functioning of the auditory system were simple, discrete and typically differed along only a single dimension. In contrast, speech signals involve very complex spectral and temporal relations. Secondly, most of the research dealing with auditory psychophysics that has accumulated over the last thirty years has been concerned with the discriminative capacities of the sensory transducer and the functioning of the peripheral auditory mechanism. In the case of speech perception, however, the relevant mechanisms are centrally located and intimately related to more general cognitive processes that involve the encoding, storage and retrieval of information in memory. Moreover, experiments in auditory psychophysics have typically focused on experimental tasks and paradigms that involve discrimination rather than identification or recognition, processes thought to be most relevant to speech perception. Thus, it is generally believed that a good deal of what has been learned from research in auditory psychophysics and general auditory perception is only marginally relevant to the study of speech perception and to an understanding of the underlying perceptual mechanisms.

Despite these obvious differences, investigators have, nevertheless, been quite interested in the differences in perception between speech and nonspeech signals. That such differences might exist was first suggested by the report on the earliest findings of categorical discrimination of speech by Liberman et al. (1957). And it was with this general goal in mind that the first so-called "nonspeech control" experiment was carried out by Liberman et al. (1961) in order to determine the basis for the apparent distinctiveness of speech sounds.

Numerous speech-nonspeech comparisons have been carried out over the years since these early studies, including several of the contributions to the present symposium. For the most part, these experiments have revealed quite similar results. Except until quite recently, performance with nonspeech control signals failed to show the same discrimination functions that were observed with the parallel set of speech signals (Cutting and Rosner, 1974; Miller et al., 1976; Pisoni, 1977). In addition, the nonspeech signals were typically responded to by subjects at levels approximating chance performance. Such differences in perception between speech and nonspeech signals have been assumed to reflect basically different modes of perception-- a "speech mode" and an "auditory mode". Despite some attempts to explain away this dichotomy, additional evidence continues to accumulate as suggested by several of the new findings summarized in the papers included in this section.

There have been, however, a number of problems involved in drawing comparisons between speech and nonspeech signals that have raised several questions about the interpretation of the results obtained in these earlier studies. First, there is the question of whether the same psychophysical properties found in the speech stimuli were indeed preserved in the nonspeech control condition. Such a criticism seems quite appropriate for the original /do/--/to/ nonspeech control stimuli which were simply inverted spectrograms as well as the well-known "chirp" and "bleat" control stimuli of Mattingly et al. (1971) that were created by removing the formant transitions and steady-states from speech context and then presenting them in isolation to subjects for discrimination. Such manipulations while nominally preserving the speech cue obviously result in a marked change in the spectral context of the signal which no doubt affects the detection and discrimination of the original formant transitions. Such criticisms have been taken into account in the more recent experiments comparing speech and nonspeech signals as summarized by Dr. Dorman and Dr. Liberman in which the stimulus conditions remain identical across different experimental manipulations. However, several additional problems still remain in making comparisons between speech and nonspeech signals. For example, subjects in these experiments rarely if ever receive any

experience or practice with the nonspeech control signals. With complex multidimensional signals it may be quite difficult for subjects to attend to the relevant attributes of the signal that distinguish it from other signals presented in the experiment. A subject's performance with these nonspeech signals may therefore be no better than chance if he/she is not attending selectively to the same specific criterial attributes that distinguish the speech stimuli. Indeed, not knowing what to listen for may force a subject to "listen" for an irrelevant or misleading property of the signal itself. Since almost all of the nonspeech experiments conducted in the past were carried out without the use of feedback to subjects, a subject may simply focus on one aspect of the stimulus on one trial and an entirely different aspect of the stimulus on the next trial.

Setting aside some of these criticisms, the question still remains whether drawing comparisons in perception between speech and nonspeech signals will yield some meaningful insights into the perceptual mechanisms deployed in processing speech. In recent years, the use of cross-language, developmental and comparative designs in speech perception research has proven to be quite useful in this regard as a way of separating out the various roles that genetic predispositions and experiential factors play in perception. For example, while it is cited with increasing frequency that chinchillas have been shown to categorize synthetic stimuli differing in VOT in a manner quite similar to human adults, little if anything is ever mentioned about the chinchilla's failure to carry out the same task with stimuli differing in the cues to place of articulation in stops, a discrimination that even young prelinguistic infants have been shown to be capable of making. Such comparative studies are therefore useful in speech perception research to the extent that they can specify the absolute lower-limits on the sensory or psychophysical processes inherent in discriminating properties of the stimuli themselves. However, they are incapable, in principle, of providing any further information about how these signals might be "interpreted" or coded within the context of the experience and history of the organism.

Cross-language and developmental designs have also been quite useful in providing new information about the role of

early experience in perceptual development and the manner in which selective modification or tuning of the perceptual system takes place. Although the linguistic experience and background of an observer was once thought to strongly control his/her discriminative capacities in a speech perception experiment, recent findings strongly suggest that the perceptual system has a good deal of plasticity for retuning and realignment even into adulthood. The extent to which control over the productive abilities remains plastic is still a topic to be explored in future research.

To what extent is it then useful to argue for the existence of different modes of perception for speech and nonspeech signals? Some investigators such as Dr. Ades and even Dr. Massaro would like to simply explain away the distinctions drawn from earlier work on the grounds of parsimony and generality. But this is a curious position to maintain as it is commonly recognized, not only in speech perception research but in other areas of perceptual psychology, that stimuli may receive differential amounts of processing or attention by the subject, that subjects may organize the interpretation of the sensory information differently under different conditions and that the sensory trace of the initial input signal may show only a faint resemblance to its final representation resulting from encoding and storage in memory. It is hard to deny that a speech signal elicits a characteristic mode of response in a human subject-- a response that is not simply the consequence of an acoustic waveform leaving a meaningless sensory trace in the auditory periphery. Such observations suggest to me that, just as in the case of "species-typical responding" observed in the behavior of numerous other organisms, the existence of a speech mode of perception is a way of capturing certain aspects of the way human observers typically respond to speech signals that are familiar to them. Such a conceptualization does not, at least in my view, commit one to the view that human listeners cannot respond to speech in other ways more closely correlated with the sensory or psychophysical attributes of the signals themselves. To explain away the speech mode, however, is to deny the fact that a certain subset of possible acoustic signals generated by the human vocal tract are used in a distinctive and quite systematic way by both

talkers and listeners to communicate by spoken language, a species-typical behavior that is restricted, as far as I know, to homo sapiens. Past experiments comparing the perception of speech and nonspeech signals have been quite useful in characterizing how the phonological systems of natural languages have, in some sense, made use of the general properties of sensory systems in selecting out the inventory of phonetic features and their acoustic correlates. The relatively small number of distinctive features and their acoustic attributes observed across a wide variety of diverse languages suggests that the distinctions between speech and nonspeech signals still remain fundamental ones setting apart research on speech perception from the study of auditory psychophysics and the field of auditory perception more generally.

References

- Cutting, J.E. and Rosner, B.S. (1974) "Categories and boundaries in speech and music", Perc. Psych. 16, 564-570.
- Lieberman, A.M., Harris, K.S., Hoffman, H.S. and Griffith, B.C. (1957) "The discrimination of speech sounds within and across phoneme boundaries", J. Exp. Psych. 54, 358-368.
- Lieberman, A.M., Harris, K.S., Kinney, J.A. and Lane, H.L. (1961) "The discrimination of relative onset time of the components of certain speech and non-speech patterns", J. Exp. Psych. 61, 379-388.
- Mattingly, I.G., Liberman, A.M., Syrdal, A.K. and Halwes, s, T.G. (1971) "Discrimination in speech and non-speech modes", Cogn. Psych. 2, 131-157.
- Miller, J.D., J.D., Wier, C.C., Pastore, R., Kelly, W.J. and Dooling R.J. (1976) "Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception", JASA 60, 410-417.
- Pisoni, D.B. (1977) "Identification and discrimination of the relative onset of two component tones: Implications for voicing perception in stops", JASA 61, 1352-1361.

SPEECH & NON-SPEECH: WHAT HAVE WE LEARNED?

Anthony E. Ades¹, Max-Planck-Gesellschaft, Projektgruppe für Psycholinguistik, Nijmegen, Netherlands.

There have been two strands of research in the speech/non-speech controversy. Firstly there are experiments where speech is compared to non-speech signals that have critical acoustic properties of speech. (See Wood, 1976, for references). This work has shown that there is no real difference: the perceptual properties of speech arise from its acoustics, not from its "speechlikeness".

This paper is concerned with the second strand, where series of speech sounds are compared to stimuli that differ along simpler dimensions like pitch and intensity. I shall summarise the arguments presented in a recent theoretical article (Ades, 1977), and extend them to other paradigms. The conclusion I shall draw is that speech/non-speech differences, as well as consonant/vowel differences, do not result from any inherent property of the sounds themselves, such as their speechlikeness, or degree of "encodedness" (Lieberman, Mattingly and Turvey, 1972), but instead depend on a property of the ensembles of stimuli used in these experiments.

This property is the range, or width of context, of the ensemble. One may think of it as the number of just-noticeable-differences across the series. This analysis is borrowed from Durlach and Braida's (1969) quantitative theory of intensity resolution. They and their colleagues, in the course of testing this theory, have obtained results for intensity that are quite analogous to results commonly obtained for speech.

There has been a consistent failure to control for the range variable when making comparisons between vowels, consonants, speech, and non-speech.

Identification and Discrimination

It all started, I think, with Miller's observation (1956) that we can discriminate far better than we can identify. Consider an intensity discrimination experiment where the subject is asked to decide if two sounds are the same or different. This can be done reliably if they are about one dB apart. Now suppose that there are 15 sounds, evenly spaced along a continuum spanning 25 dB. About

(1) This paper was prepared while the author held a Fellowship under the Royal Society European Science Exchange Programme.

25 discriminations could be made in this space. But when the subject is asked to label the stimuli with a number between 1 and 15 (give as much practice and feedback as possible), only seven plus or minus two categories can be used accurately. The subject is distinguishing between adjacent stimuli in identification less than half as well as (s)he would in discrimination. How can this be? After all, sensitivity to acoustic signals and to differences between them must be the same in both situations.

The answer must lie in the memory requirements. In discrimination there are two or more stimuli: the subject must store their sensory traces, compare them (perhaps by subtraction), and pronounce on the difference if any. Call this the trace mode. In the identification case, a single sensory trace must be compared to some representation of the entire stimulus series. Where, the subject must decide, does this stimulus fit, given all the others I have heard. This is the context coding mode. Presumably, the representation of the series is not in the form of traces, but is in some verbal or numerical code.

Now consider what happens to identification when the range of the ensemble is increased from 25 to 50 dB. A reasonable guess is that "accuracy", defined as the ability to place the current stimulus in context, expressed as a percentage of the size of the context, will remain constant. Of course, the absolute size of errors, expressed in j.n.d. terms, will now be larger. An archer who remains a constant 3 degrees off centre will show a larger absolute error at 50 yards than at 25.

So far, then, we assume that in discrimination (in its ideal form), the only factor affecting performance is the noisiness of the representation of the acoustic traces. In identification there will be trace noise too, but there will also be context coding noise when the subject attempts to locate a stimulus in its context. This will increase as the range increases.

The critical prediction is that as long as range is small, identification performance will be as good as discrimination. For, context noise will be minimal, and trace noise will be the only determinant in both tasks.

The theorising above is an informal statement of Durlach and Braida's (1969) quantitative theory for intensity resolution. The

above prediction, that as range decreases, identification improves, and finally approximates discrimination, was confirmed by Pynn, Braida, and Durlach (1972), for stimuli differing in intensity.

And now to speech. The classic result (Studdert-Kennedy et al., 1970) is that for series of consonant - vowel stimuli, discrimination is scarcely better than identification. Given Miller's paper on how identification is relatively weak in non speech, it was natural to see the speech results as evidence for a speech-specific mode of processing. The alternative I propose is simply that CV series are not unusual by virtue of being speech, but simply have relatively small ranges. I have shown elsewhere (Ades, 1977) that the best estimates of the range of CV series make them comparable in j.n.d. terms to the small ranges used by Durlach, Braida, and their colleagues for intensity resolution experiments.

Typically, a series of synthetic speech sounds from /ba/ to /da/, or from /ba/ to /pa/, spans between 3 and 5 j.n.d.s. A series of vowels, on the other hand must stretch across about 10 j.n.d.'s to reach from one category to another. We thus expect that discrimination on vowels will far exceed what would be predicted from identification. This has been consistently found. Generally, though, it has been interpreted to mean that vowels are somehow less "speech-like" than consonants (as if one could have stop consonants without vowels!). It should be clear that I am trying to replace this rather mystical theorising with the idea that speech, non-speech, vowels and consonants are all the same. The observed differences are due to the range variable. The number of j.n.d.'s across the series, can be used as a stimulus-free approximation of the size of the range.

More complex experiments

In certain cases, the range has an effect in discrimination experiments, not just in identification. This is because certain variations in the task parameters may make it profitable for the subject to operate in the context mode, i.e. to do identification: as, for instance, when the procedure adds noise or interference to the sensory trace mode. We can predict that any manipulation that makes comparison of traces harder will only worsen performance if the range is large! For, if the range is small, the subject can escape the trace noise, slip into the context-coding mode and not

suffer too much from context noise.

In these cases it is important how discrimination is tested: if the pair to be discriminated randomly changes from trial to trial, then the effective range is the range of the entire series. But if the same pair is tested many times before another part of the series is tested, the effective range will obviously be very small. It turns out that in speech research the "roving level" method is always used. Thus procedures that cause trace comparison to be harder, such as increasing the time interval between the two stimuli, or by forcing the subject to compare three traces at a time rather than two, such procedures will, in roving level testing, make the range variable critical.

Experiments of this type have been done with vowels and consonants (Pisoni, 1973, 1975). As we predict from the Durlach and Braida model, manipulations that worsen discrimination have a stronger effect on vowels than on consonants, because, according to the range hypothesis, the small range of consonants makes escape into context-coding possible without running into context memory noise. In intensity resolution, Berliner and Durlach (1973) have shown that increased time delay between stimuli to be discriminated worsens resolution only if the range is large.

The "anchor" effect and RT Experiments

The same ideas can be applied to other paradigms where speech and non-speech have been contrasted. In the two areas that follow I confess to being less certain of my argument, because I do not know of research where the range variable has been systematically studied.

Firstly, the "anchor effect". A series of sounds varying in pitch is constructed and the subject asked to identify them as "High" or "Low". If an endpoint stimulus (the anchor), say the highest pitched one, is presented two or three times as often as the others, the entire identification curve is shifted towards to the anchor. However, such shifts do not occur in stop-consonant series (Sawusch and Pisoni, 1973; Simon and Studdert-Kennedy, 1978). Again, we might expect that the different ranges of pitch and consonant series are involved. We may assume a 5 j.n.d. range for the speech series. The pitch series went from 114 Hz to 150 Hz: assuming a difference limen of 0.5 Hz for pitch (Klatt, 1973), this

series would span over 50 j.n.d.s. Both Simon et al., and Sawusch et al. (1974) also found a strong anchor effect in a series varying in intensity. This covered 18 dB in one experiment and 24 in the other, about 20 j.n.d.s.

Certainly, then, the range differences between the speech and non speech series were marked. But why should the range determine the anchor effect? I have no formal answer to this, but it is clear that anchor effects cannot be located in the trace mode. Also, Berliner, Durlach and Braida (1977) have shown that the "edge effect", whereby resolution in identification is better at the ends of a continuum than in the middle, and which is identified in their model as a perceptual anchoring effect in the context coding mode, is enhanced by increased range.

A second paradigm is a Reaction Time task where the subject must press one of two buttons depending on whether the stimulus is /ba/ or /da/, or whether it has high or low pitch. The point here is that if the subject is responding to the speech distinction, irrelevant variation in pitch slows the RT. However, irrelevant variation in place of articulation has a much smaller effect on RT to the pitch distinction (Day and Wood, 1972). Wood (1973) also showed that there was mutual interference between pitch and intensity, and also between place of articulation and voicing. This was interpreted as revealing two separate systems: such that there was interference within each, speech with speech, non-speech with non-speech; but no interference between.

The alternative is that both pitch and intensity discriminations are easy, while both place and voicing are harder. Interference will occur if the irrelevant variation is as salient or more salient than the distinction being tested. The situation where interference is least is precisely the one where the discrimination (pitch) is much more salient than the interfering dimension (place).

Finally, let me add that the point I have been trying to make for discriminations vs identification, anchor effects, and RT experiments has already been forcefully made for experiments on the Precategorical Acoustic Store (PAS), and on the hemispheric lateralisation of speech. The fact that sets of stop-consonant-vowel syllables produce no recency effect in PAS, whereas sets of vowels do, has been taken to mean that consonants and vowels are differen-

tially "encoded" (Liberman et al., 1972). But Darwin and Baddeley (1974) have shown that the vowel/consonant distinction here is irrelevant: what controls the recency effect is, again, the discriminability of the items within the ensemble. Similarly, the same factor is critical in determining the degree of hemispheric lateralisation for vowels (Godfrey, 1974).

Conclusions

At the very least it must be conceded that explorations of speech/non-speech and vowel/consonant differences might be meaningless unless factors corresponding to discriminability across the stimulus ensemble are controlled. It is obvious that the range variable is all-important in the experiments briefly reviewed here. In addition, once range is controlled for, a single unified theory for all stimuli seems well within reach. And this is surely preferable to one theory for non-speech, a second theory for consonants, (and an in-between theory for vowels).

Whether or not the above proposals are correct, the entire speech/non-speech issue seems to have acquired a life of its own, which it fights for against all odds. However, according to the views expressed here, it has taught us very little, and has simply served to direct out attention from the real problems of speech perception, exemplified for example in automatic recognition (Klatt, 1977, for a review), where the psychological contribution remains slight and engineering solutions prevail.

References

- Ades, A. E. (1977): "Vowels, Consonants, Speech, and Nonspeech", *Psych. Rev.* 84, 524-530.
- Berliner, J. E., and N. I. Durlach (1973): "Intensity Perception IV: Resolution in Roving Level Discrimination", *JASA*, 53, 1270-87.
- Berliner, J. E., N. I. Durlach, and L. D. Braida (1977): "Intensity Perception VII. Further Data on Roving Level Discrimination and the Resolution and Bias Edge Effects", *JASA*, 61, 1577-85.
- Darwin, C. D., and A. D. Baddeley (1974): "Acoustic Memory and the Perception of Speech", *Cogn. Psych.* 6, 41-60.
- Day, R. S., and C. C. Wood (1972): "Interaction between Linguistic and Nonlinguistic Processing", *JASA*, 51, 79(A).
- Durlach, N. I., and L. D. Braida (1969): "Intensity Perception I. Preliminary Theory of Intensity Resolution", *JASA*, 46, 372-83.

- Godfrey, J. J. (1974): "Perceptual Difficulty and the Right-Ear Advantage for Vowels". Brain and Language, 4, 323-36.
- Klatt, D. H. (1973): "Discrimination of Fundamental Frequency Contours in Synthetic Speech: Implications for Models of Pitch Perception", JASA, 53, 8-16.
- Klatt, D. H. (1977): "Review of the ARPA Speech Understanding Project", JASA, 62, 1345-66.
- Liberman, A. M., I. G. Mattingly, and M. T. Turvey (1972): "Language Codes and Memory Codes". In A. W. Melton and E. Martin (Eds.) Coding Processes in Human Memory, Washington, D. C.: Winston.
- Miller, G. A. (1956): "The Magical Number Seven, Plus or Minus Two: Some Limits on our capacity for Processing Information", Psych. Rev., 63, 81-97.
- Pisoni, D. B. (1973): "Auditory and Phonetic Codes in the Discrimination of Consonants and Vowels", Perc. Psych., 13, 253-60.
- Pisoni, D. B. (1975): "Auditory Short-Term Memory and Vowel Perception", Memory and Cognition, 3, 7-18.
- Pynn, C. T., L. D. Braida, and N. I. Durlach (1972): "Intensity Perception III. Resolution in Small-Range Identification", JASA, 51, 559-66.
- Sawusch, J. R., and D. B. Pisoni (1973): "Category Boundaries for Speech and Nonspeech Sounds", JASA, 54, 76(A).
- Sawusch, J. R., D. B. Pisoni and J. E. Cutting (1974): "Category Boundaries for Linguistic and Nonlinguistic Dimensions of the Same Stimuli", JASA, 55, S55(A).
- Simon, H. J., and M. Studdert-Kennedy (1978): "Selective Anchoring and Adaption of Phonetic and Nonphonetic Continua", JASA, 64, 1338-57.
- Studdert-Kennedy, M., A. M. Liberman, K. S. Harris, and F. S. Cooper (1970): "Motor Theory of Speech Perception: A Reply to Lane's Critical Review", Psych. Rev., 77, 234-49.
- Wood, C. C. (1973): "Levels of Processing in Speech Perception. Neurophysiological and Information Processing Analyses". Unpublished Doctoral Dissertation, Yale University.
- Wood, C. C. (1976): "Discriminability, Response Bias, and Phoneme Categories in Discrimination of Voice Onset Time", JASA, 60, 1381-89.

SOME PSYCHOACOUSTIC FACTORS IN PHONETIC ANALYSIS

Pierre L. Divenyi, Veterans Administration Medical Center, Martinez, California, 94553.

From an ethological point of view, speech represents a complex acoustic stimulus that has the greatest survival value for man. Physically speaking, speech is complex in two ways: its spectral composition, over any epoch of arbitrary length, is extremely rich, and this spectral composition is continuously varying over time. The information density represented by the speech signal is enormous; yet, the human auditory system, despite its limited capacity, is able to receive and decode such a complex signal with remarkable efficiency. The desire to provide a reasonable explanation for such efficiency, as well as the need for descriptive data on the perceptual processes that permit reception and decoding of speech, provided much of the motivation behind the greatest part of the speech perception research accomplished to date. The emerging body of experimental findings, in turn, has constituted the background for a number of theories and models of speech perception. The leitmotiv of many of these theories, including some major contemporary ones, is that speech represents a special acoustic signal that must be handled by the auditory system in a special way (=speech mode), involving special processes and mechanisms (=phonetic feature detectors, etc.). While the special nature of speech and speech perception processes can hardly be disputed (because of their aforementioned high survival value), some recent results demonstrating speech discrimination by young infants and animals have established the need for an alternative theoretical approach-- one that would take into account, at least to some extent, some "wired-in" properties of the auditory mechanisms. The purpose of the present paper is to invoke some basic properties of the human auditory system and to reflect on the consequences of these properties for the phonetic analysis of the speech signal.

Psychophysical reality of the speech signal

Classical psychoacoustics research and classical speech perception research have progressed on traditionally separate (and not always parallel) paths. The reasons for this divorce, considered by some cynics as permanent until quite recently, were numerous, one of them being the overwhelming concern of psychoacousticians with simple acoustic signals and peripheral auditory processes. How-

ever, for the last couple of decades, the situation has gradually changed: availability of sophisticated stimulus control, the growing popularity of a systems approach to perceptual problems, and interdisciplinary orientation of an increasing number of researchers have signaled the beginnings of a (hopefully) new era. Indeed, psychoacoustics appears to be no longer afraid of spectrally and/or temporally complex sound patterns and researchers seem to address with greater freedom issues involving more central processes. Thus, it has become possible to take a fresh look upon the speech signal as a stimulus to the auditory system, and to interpret its perception in terms of a certain number of discrete psychoacoustic processes. For reasons of economy, only a few major ones will be discussed here.

Peripheral analysis and time-frequency trade. Peripheral analysis of auditory signals operates under a constraint not unlike Heisenberg's Uncertainty Principle, as defined for elementary particle physics. According to this principle, in any given system frequency resolution (Δf) can be traded for temporal resolution (Δt) and vice versa, such that their product $\Delta f \Delta t$ remains constant. In the ear, such a relation is generally true only within certain limits (McGill, 1968); spectral resolution is limited by the Critical Bands (roughly 1/4 to 1/3 octaves in width; Zwicker et al., 1957) and temporal resolution by the ear's "time window" (a time constant of roughly 8 msec; Penner, 1978). Within these limits, however, this principle predicts that, to increase resolution in the spectral domain, temporal resolution must be sacrificed, and vice versa (Ronken, 1971). The validity of this prediction is proved by experimental results: discrimination of the frequency of pure tones deteriorates as their duration decreases (Moore, 1973) and, conversely, perception of the fine temporal structure of the stimulus is possible only for wide-band signals (Green, 1971).

Thus, the length of the effective time window and the width of the effective internal filter continuously adapt themselves to the spectral-temporal characteristics of the stimulus. The outcome of such an analysis will be a sequence of "neural spectra" (Klatt, 1978) or "central spectra" (de Boer, 1977) -- a series of quasi-stationary auditory events of variable duration. The temporal constraint signifies that peripheral analysis of acoustic (speech or non-speech) signals cannot be extended beyond the duration of these auditory events.

Pitch perception. According to contemporary theories (Plomp, 1975), pitch of complex signals is extracted by periodicity analysis of the internal spectrum (i.e., by taking its Fourier transform). Thus, any complex signal gives rise to two different pitch experiences: a "spectral pitch" (=formant analysis) and a "virtual pitch" (or low pitch or residue pitch [=periodicity analysis]), the former being a prerequisite for the latter. The existence region of virtual pitch is limited to pitch periods not shorter than about 2 msec (< 500 Hz); the degree of its salience is a composite function of the spectral region (formant region), the serial number and the relative intensity of the component harmonics, and the periodicity rate itself (Ritsma, 1962). In complex signals consisting of several consecutive harmonics virtual pitch is determined by the eight lowest harmonics, especially those around the third (Houtgast, 1974, 264), but, interestingly, the fundamental is not dominant.

Virtual pitch is not an absolute concept: it reflects a statistical approximation to a periodicity that derives from the ensemble of peaks in the internal spectrum (de Boer, 1977). It has also been proposed (Terhardt, 1974) that virtual pitch actually represents a Gestalt property of complex sounds -- a property that is as much a result of learning as that of purely sensory processes. Such a hypothesis helps account for some systematic pitch shift phenomena that are otherwise difficult to interpret.

Temporal organization. Since peripheral analysis is limited to short temporal intervals, the sequences of "neural spectra" which temporally-complex signals generate must be organized into perceptually meaningful units by some higher-level auditory center(s). Such a perceptual organization in time obeys rules that are reminiscent of the Gestalt principles that govern the perception of visual figures in space (e.g., law of closure, law of proximity, etc; Koffka, 1935) and, ultimately, leads to the percept of an auditory pattern (Divenyi and Hirsh, 1978). Among the general rules of auditory pattern perception there is one of primary importance: two successive auditory events can be optimally resolved in time only if they occur in identical spectral bands. For example, auditory discrimination of short (10-30 msec) intervals defined by the onsets of two brief tones gradually deteriorates when the two tone frequencies become increasingly different (Divenyi and Sachs, 1978). Similarly, recognition of the temporal order of successive tones remains accurate only as long as all tone frequencies are within the same nar-

row band -- otherwise the sequence breaks into separate "auditory streams" (Bregman and Campbell, 1971).

The concept of "listening bands". The three above mentioned limitations, i.e., trade-off of time resolution - frequency resolution, limits of periodicity analysis, and restriction of accurate temporal organization to auditory events within the same narrow spectral band, are generally valid for the processing of any auditory signal, simple or complex. However, since speech constitutes an auditory stimulus in which the spectral information is generally distributed over several bands (specific to a given phonetic unit), its processing will be further complicated by yet another limitation: the auditory system is unable to simultaneously monitor several bands without loss of information (Green, 1961). The consequence of such a limitation is that auditory processing along various acoustic dimensions will be degraded by frequency uncertainty, i.e., by leaving the listener in doubt as to the frequency region in which the forthcoming auditory event is to appear. For example, frequency uncertainty will degrade detection (Creelman, 1972) and frequency discrimination (Watson, 1976) of a pure tone, as well as recognition of temporal-order patterns of several successive tones (Divenyi and Hirsh, 1978).

In order to overcome the effect of frequency uncertainty, the auditory system tends to spontaneously "tune" its focus of listening to the narrow band at or around the input frequency; it will usually remain focused at this listening band in the absence of any stimulus for at least several seconds (Johnson, 1978). Thus, at any given time, the auditory system's choice of a listening band is determined by the frequency characteristics of the last input. One of the possible reasons for the detrimental effect of frequency uncertainty is that shifting the listening focus from one band to another seems to take time (Divenyi and Hirsh, 1972). Moreover, attending to more than one spectral band at once will also degrade listening efficiency (Swets, 1963) -- the information processing capacity of the ear is, indeed, quite limited. The surprising finding is that the listener's knowledge with regard to the frequency of the forthcoming stimulus is not sufficient to completely eliminate the frequency uncertainty effect: to tune the listening band to a new region some sound (i.e., a cue) must occur (Johnson, 1978).

The locus of the tuning mechanism most probably lies above the auditory periphery: contralateral cues, too, have been found to be

effective in establishing the listening bands (Gilliom et al., 1979)

Relevance to speech perception

The question of great interest to many is how a system having the properties described above is likely to behave when confronted with a speech signal. While a great deal more experimental data than what we have to date are needed to answer this question (even in a marginally acceptable manner), it is nonetheless possible to give a cursory outline of the effects of auditory processing on speech sounds. Again, because of space limitations, the picture presented here will be sketchy and less than exhaustive.

Segmentation. As a direct result of the time resolution - frequency resolution trade-off, any complex signal in which narrow-band and wide-band portions alternate will be automatically segmented at a peripheral level. Since, in speech, transitions from wide-band to narrow-band acoustic segments (and vice versa) roughly correspond to phonetic segment dividers, each of these transitions (smoothed by the ear's time window function) will produce marker signals at the auditory periphery. Thus, the series of auditory events (= "neural spectra") which some higher-level centers will organize into perceptual units will actually be a succession of phonetically meaningful elements.

Speaker invariance. The mutual interdependence of waveform periodicity, spectrum of complex sounds, salience of virtual pitch, and salience of spectral pitch can account for much of the formant frequency - fundamental frequency relations observed in vowel production and perception (Fujisaki and Kawashima, 1968). Since vowels (=quasi-steady-state sounds) are analyzed in a narrow-band mode, relatively small spectral variations may be detected by the auditory system. Such a large degree of sensitivity may provide the explanation underlying the notion that vowel perception is "continuous" rather than "categorical".

Categorical perception and selective adaptation. In CV syllables, especially in stop-vowel pairs, the initial consonant is a wide-band transient; therefore, nothing compels the auditory system to tune the listening band to any particular position of the spectrum. The relative freedom of tuning that derives from wide-band stimuli enables the auditory system to select a frequency region to which it will spontaneously direct its focus before the onset of the CV sound. Such strategies may possibly originate in learning:

category boundaries that characterize certain features are known to be language-bound. However, strategies for positioning the listening band are by no means absolute: a sound of different spectral-temporal characteristics (speech or non-speech, see Samuel and Newport, 1979) presented prior to the CV stimulus could serve as a cue (Johnson, 1978) and make the auditory system choose a different listening band. Thus, selective adaptation effects could be re-interpreted in terms of pre-cueing and listening bands.

Such an interpretation is quite straightforward when one looks upon category boundary shifts observed for the feature of place-of-articulation in adaptation experiments: the acoustic basis for this feature is almost exclusively spectral. Explanation of boundary shifts of the voiced-voiceless category, a predominantly temporal feature, is somewhat more complex. Since temporal organization of acoustic events heavily depends on temporal cues contained in some narrow band, perception of the feature of voicing will be a function of the discriminability of voice-onset-time inside one (or several) narrow spectral region(s). However, when a brief auditory time interval is marked by a pair of sounds of identical spectral composition, temporal masking (forward or backward) of one marker by the other could decrease the discriminability of the interval (Divenyi and Sachs, 1978). Because the relative energy of the consonant and the vowel varies from one band to another (thereby also causing the amount of temporal masking to vary), the choice of the monitored band will be critical in determining the VOT boundary. Thus, tuning the listening band to different spectral regions will result in different voicing boundaries. An adaptor stimulus (by virtue of its potential role as a cue), therefore, may alter the natural position of the listening band for a given CV syllable, thereby producing a shift in the category boundary. It is conceivable that perceptual-productive acquisition of different phonetic patterns could also be associated with different spectral positions that the listening band will spontaneously occupy; thus, the present theory is consistent with the language-dependent nature of voicing category boundaries.

Time invariance implies that the relative duration of certain phonetic segments is irrelevant. Experiments on the perception of non-speech sound sequences (Watson, 1976; Divenyi and Hirsh, 1978) have shown that the emergence of an auditory pattern (at least within certain limits) does not depend on the absolute duration of the

components. Thus, it follows that the rate at which the speech segments ("neural spectra") of the speech sounds occur will not change the "figural properties" of the patterns.

Conclusion: Whither phonetic analysis?

When attempting to examine speech processing on the auditory level, one finds that the product of auditory analysis possesses several characteristics that are customarily thought to belong to the realm of phonetic analysis (feature analysis, etc.). While it is readily acknowledged here that many crucial experiments needed to prove (or disprove) critical points have not yet been performed, and that straight extrapolation of non-speech auditory data to speech-bound processes may often be risky, we feel, nevertheless, that auditory analysis of the speech signal well exceeds the limits imposed on it by several widely accepted theories. The view that phonetic analysis may not be an indispensable stage in speech processing is concordant with the opinion expressed in some studies on the perception of speech by man (Ades, 1976) or the recognition of speech by machine (Klatt, 1978). An alternative view, one that we would like to propose herewith, is that speech perception may be regarded as a special class of auditory pattern perception -- special only because we have learned these patterns so well.

References

- Ades, A.E. (1976): "Adapting the property detectors for speech perception", in New approaches to language mechanisms, R.J. Wales and E. Walker (eds.), 55-108, Amsterdam: North Holland.
- de Boer, E. (1977): "Pitch theories unified", in Psychophysics and physiology of hearing, E.F. Evans and J.P. Wilson (eds), 323-334, London: Academic.
- Bregman, A.S. and J.L. Campbell (1971): "Primary auditory stream segregation and perception of order in rapid sequences of tones", JEP 89, 244-249.
- Creelman, C.D. (1972): "Detecting signals of uncertain frequency: Analysis by individual alternative signals", JASA 52, 167.
- Divenyi, P.L. and I.J. Hirsh (1972): "Discrimination of the silent gap in two-tone sequences of different frequencies", JASA 51, 138.
- Divenyi, P.L. and I.J. Hirsh (1978): "Some figural properties of auditory patterns", JASA 64, 1369-1385.
- Divenyi, P.L. and R.M. Sachs (1978): "Discrimination of time intervals bounded by tone bursts", Perc. Psych. 24, 429-436.
- Fujisaki, H. and T. Kawashima (1968): "The roles of pitch and higher formants in the perception of vowels", IEEE AEA AU-16, 73-77.
- Gilliom, J., D.W. Taylor and C. Cline (1979): "Timing constraints

- for effective cueing in the detection of sinusoids of uncertain frequency", Perc. Psych. 25 (in press).
- Green, D.M. (1961): "Detection of auditory sinusoids of uncertain frequency", JASA 33, 897-903.
- Green, D.M. (1971): "Temporal auditory acuity", Psych. Rev. 78, 540-551.
- Houtgast, T. (1974): "Masking patterns and lateral inhibition", in Facts and models in hearing, E. Zwicker and E. Terhardt (eds.), 258-265, Berlin: Springer.
- Johnson, D.M. (1978): "Attentional factors in the detection of uncertain auditory signals", Unpubl. Doct. Dissert. Univ. Calif. Berkeley.
- Klatt, D.H. (1978): "Speech perception: A model of acoustic-phonetic analysis and lexical access", in Perception and production of fluent speech, R.A. Cole (ed), Hillsdale (N.J.): Erlbaum.
- Koffka, K. (1935): Principles of Gestalt psychology, New York: Harcourt Brace.
- McGill, W.J. (1968): "Polynomial psychometric functions in audition", J. Math. Psych. 5, 369-376.
- Moore, B.C.J. (1973): "Frequency difference limens for short-duration tones", JASA 54, 610-619.
- Penner, M.J. (1978): "A power-law transformation resulting in a class of short-term integrators that produce time-intensity trades for noise bursts", JASA 63, 195-201.
- Plomp, R. (1975): "Auditory psychophysics", Ann. Rev. Psych. 26, 207-232.
- Ritsma, R.J. (1962): "Existence region of the tonal residue", JASA 34, 1224-1229.
- Ronken, D.A. (1971): "Some effects of bandwidth-duration constraints on frequency discrimination", JASA 49, 1232-1242.
- Samuel, A.G. and E.L. Newport (1979): "Adaptation of speech by non-speech: Evidence for complex acoustic cue detectors", JEP HPP 5 (in press).
- Swets, J.A. (1963): "Central factors in auditory frequency selectivity", Psych. Bull. 60, 429-440.
- Terhardt, E. (1974): "Pitch, consonance, and harmony", JASA 55, 1061-1069.
- Watson, C.S. (1976): "Factors in the discrimination of word-length auditory patterns", in Hearing and Davis: Essays honoring Hal-lowell Davis; S.K. Hirsh, D.E. Eldredge, I.J. Hirsh, and S.R. Siverman (eds.), 175-189, St. Louis: Washington University Press.
- Zwicker, E., G. Flottorp and S. S. Stevens (1957). "Critical band width in loudness summation", JASA 29, 548-557.

ON THE IDENTIFICATION OF SINE-WAVE ANALOGUES OF CV SYLLABLES¹

Michael F. Dorman,² Haskins Laboratories, 270 Crown Street,
New Haven, Connecticut 06510, United States of America

In order to answer the question - Do infants perceive speech phonetically? - the stimulus continuum presented to the subjects must have phonetic category boundaries which are clearly dissociated from auditory category boundaries. For, if the two boundaries coincide, then the subjects' basis for response can not be determined. This situation, in the view of several authors, characterizes the identification of categories along the voice-onset-time (VOT) continuum. For example, Pisoni (1977) suggests that the auditory categories of simultaneous and nonsimultaneous onset could underlie infants' discrimination along the VOT continuum.

In the present series of experiments our aim was to determine whether auditory categories may also underlie infants' ability to discriminate between stop consonants which differ in place of articulation. An examination of the stimuli used in Eimas' (1974) and Miller and Morse's (1976) studies of infant place discrimination suggests a possible psychoacoustic basis for the discrimination between [bae] and [dae] - i.e., the discrimination could be based on the difference between frequency change and no frequency change in the second and third formant transitions. While the outcomes of the two studies lend little support to this position, we felt, nevertheless, that it would be important for future research to assess whether auditory categories generally coincide with phonetic categories along a continuum of F_2 and F_3 change.

The procedure used in our experiments was to present adults with consonant-vowel (CV) syllables synthesized with formant structure and CV analogues synthesized with frequency and amplitude modulated sine waves. Our rationale for this approach was that if listeners placed category boundaries at the same place along both the speech and nonspeech continua, then we should believe that, for these stimuli at least, the phonetic category boundaries coincide with acoustic category boundaries. These stimuli, of course, would be inappropriate for use with infants. If, on the other hand,

-
- (1) This research was a collaborative effort among Dr. Peter Bailey, Dr. Quentin Summerfield and myself.
 - (2) Also, Arizona State University, Tempe, Arizona 85281, United States of America.

the two boundaries did not coincide, then the speech stimuli could well prove probative in studies of infant speech perception.

The stimuli for our first experiment were a [bo-do] continuum and a [be-de] continuum (see Figure 1). The first and third formants in both continua were identical - only the second formant differed between the two. The parameter values were selected so that both continua would be physically symmetrical but phonetically asymmetrical. We intended the phoneme boundary along the [bo-do] continuum to be associated with a falling transition so that the majority of the stimuli would be heard as [bo]. In contrast, we intended the phoneme boundary along the [be-de] continuum to be associated with rising transitions so that the majority of the stimuli would be heard as [de]. In this way we intended to dissociate phonetic boundaries from auditory boundaries that may accompany flat as opposed to rising or falling transitions, or from auditory boundaries that might simply coincide with the center of the stimulus range.

To generate identification functions for these stimuli, we presented the stimuli to our listeners in an AXB format. On each trial three stimuli were presented; the first and third members of the triad were the end points of the continuum, the second member was a stimulus drawn randomly from that continuum. The task of the listeners was to indicate whether the second stimulus was more like the first or more like the last member of the triad. We chose this task to avoid a problem usually associated with the absolute identification of nonspeech stimuli - that listeners have more difficulty attaching category labels to the nonspeech stimuli than to the speech stimuli. By presenting the end points of the stimulus continuum on each trial in both the speech and nonspeech conditions, we hoped to make the identification task equally difficult in both conditions.

Turning now to the result of our first experiment, we see in Figure 2 the identification function for the speech signals. As predicted there were more [b]-like responses for the [bo-do] continuum than for the [be-de] continuum. However, the difference between the locations of the phoneme boundaries fell short of significance. In contrast to the largely asymmetric identification functions shown for the formant stimuli, the identification functions for the sine-wave analogues, shown in Figure 3, coincide

throughout their range. We would conclude from this outcome that at least the [bo-do] boundary does not coincide with an acoustic category boundary. There are, however, two possible interpretations of this outcome: the asymmetrical categorization of the formant continua could either be correlated with the way the stimuli are heard - as speech or nonspeech - or may simply be correlated with the different spectral properties of the formant and sine-wave stimuli. To rule out the latter possibility we would like the same physical signal to be heard as speech-like in one context and as nonspeech in another. If the category boundaries differed in this instance then it certainly could not be argued that spectral differences account for the outcome. Fortunately, the sine-wave stimuli used in our experiment fit this requirement nicely. After we instructed our listeners as to the nature of the sine-wave stimuli, they readily agreed that the stimuli could be heard as stop initiated.

The outcome of this experiment (when the sine waves were heard as speech) is shown in Figure 4. The pattern of results is clearly very different from that when the sine waves were heard as nonspeech. Here, the two functions no longer overlap. As with the formant stimuli, the majority of the [bo-do] analogues were heard as [b]-like. Moreover, the category boundaries along the two continua differed significantly. It is clear that the pattern of results obtained when the sine waves were heard as speech-like is more akin to that obtained for the formant stimuli than for that obtained when the sine waves were heard as nonspeech. It appears, then, that the difference between the speech and nonspeech conditions was not due to the spectral differences as such, but, rather, was due to the way in which the stimuli were heard.

To assess the reliability of our first experiment we conducted a second experiment. For this experiment we synthesized a single [ba-da] continuum and a corresponding nonspeech analogue with sine waves. The speech continuum was more natural sounding than either of those used in our first experiment and, perhaps as a consequence, many listeners heard the sine-wave analogues as speech-like without prompting. Thus, we were able to divide our subjects into two groups on the basis of their perception of the sine-wave stimuli.

The identification function for the subjects who heard the sine waves as speech is shown in Figure 5 along with the identifica-

tion function for the formant stimuli. The two functions are quite similar and two phoneme boundaries fall to the right of the mid-point of the stimulus continuum. In contrast, the identification function for the subjects who heard the sine waves as nonspeech appears quite different from that generated in response to the formant stimuli (see Figure 6). This difference is reflected in the significantly flatter slope of the nonspeech function and fewer [b]-like responses to the nonspeech stimuli.

Summary

Earlier in this paper we raised the question of whether auditory category boundaries generally coincide with phonetic category boundaries along F_2 - F_3 continua. Unfortunately, our results provide an equivocal answer; the [bo-do] boundary in our first experiment did not seem to coincide with the auditory boundary, but the phonetic and auditory boundaries in the second experiment were uncomfortably close. Nevertheless, we have gained a significant purchase on a methodology that will allow us to dissociate auditory and phonetic boundaries. We see, then, an opportunity to construct continua which will be of use in the study of the ontogeny of phonetic perception.

Moreover, we see quite clearly that the perceptual system categorizes sine-wave stimuli as a function of how they are heard: when heard as speech they are categorized like formant stimuli; when heard as nonspeech they are categorized differently. We should wonder then what mechanisms underlie this changing percept of an unchanging stimulus. The nature of those mechanisms will be, I believe, the topic of Dr. Bailey's and Dr. Summerfield's paper.

References

- Eimas, P.D. (1974): "Auditory and linguistic processing of cues for place of articulation by infants", *Perc.Psych.* 16, 513-521.
- Miller, C.L. and P.A. Morse (1976): "The 'Heart' of categorical speech discrimination in young infants", *JSHR* 19, 578-589.
- Pisoni, D.B. (1977): "Identification and discrimination of the relative onset times of two component tones: Implications for voicing perception in stops", *JASA* 61, 1352-1361.

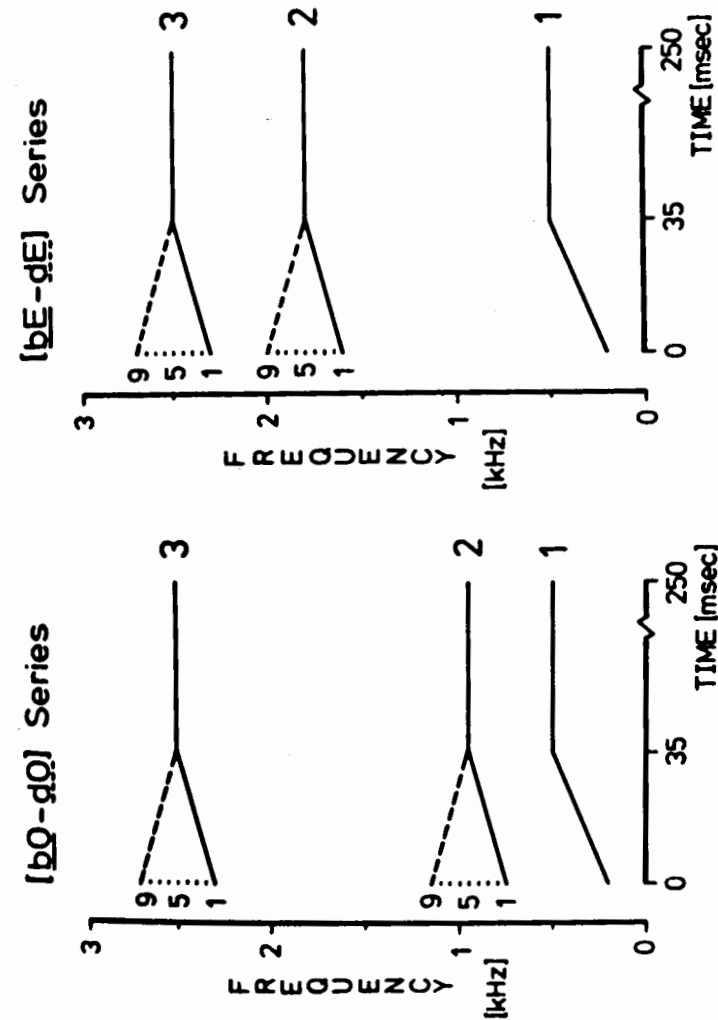


Figure 1
Stimuli for first experiment

FORMANTS heard as SPEECH

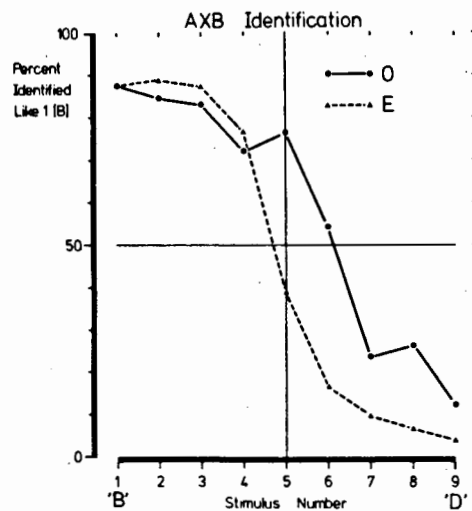


Figure 2

SINE-WAVES heard as NON-SPEECH

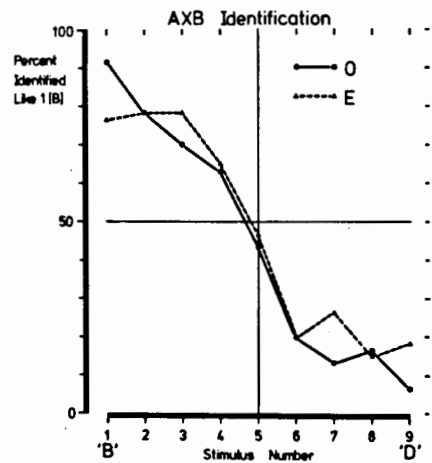


Figure 3

SINE-WAVES heard as SPEECH

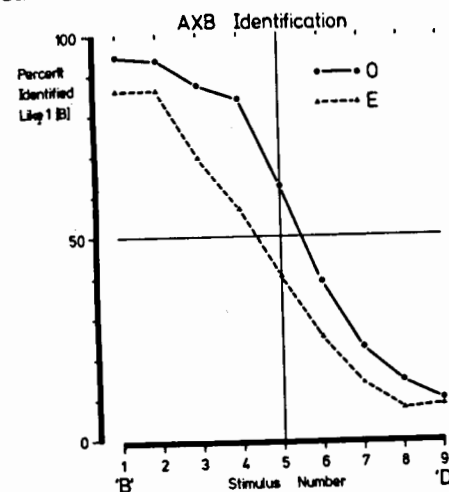


Figure 4

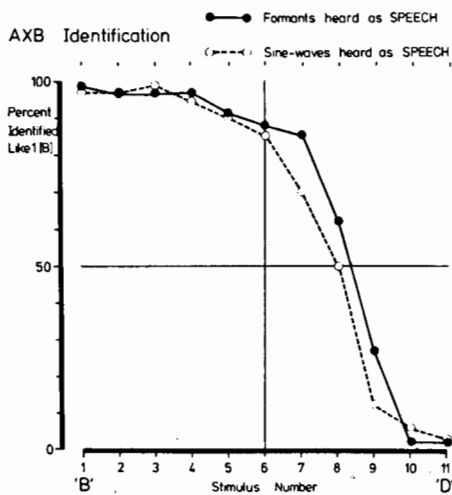


Figure 5

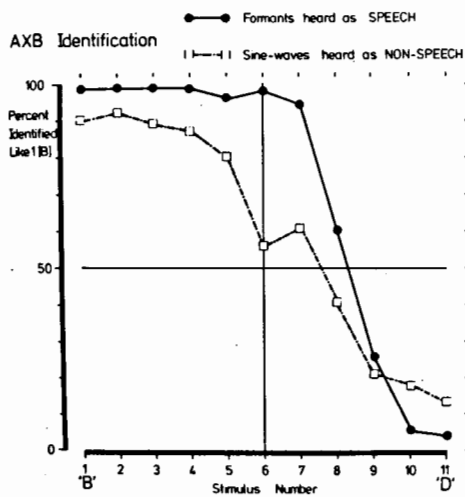


Figure 6

PERCEPTUAL LEARNING OF MIRROR-IMAGE ACOUSTIC PATTERNS¹

M.E. Grunke and D.B. Pisoni², Psychology Department, Indiana University, Bloomington, Indiana 47401 U.S.A.

The phonetic units of spoken language are often mapped in a one-to-many fashion onto their acoustic representations in speech. As a consequence, a child acquiring language must in some way be able to recognize a variety of different acoustic signals as members of the same phonetic category. For example, the /b/ in the syllable /ba/ is characterized by rapidly rising formant transitions into the following vowel, whereas the same consonant in the syllable /ab/ is characterized by rapidly falling formant transitions.

In view of the lack of one-to-one correspondence between phonemes and their acoustical representations in speech, it would seem advantageous for a child learning language if the various acoustical forms of a particular phoneme were related to each other perceptually. In fact, the acoustical representations of stop consonants in initial and final position, although physically different, are related: stop consonants in final syllable position are roughly the mirror image in time of their counterparts in initial position. But are mirror-image acoustic patterns inherently related perceptually for the listener?

The issue of the perceptual relatedness of mirror image acoustic patterns was addressed recently in a series of experiments by Klatt & Shattuck (1975) & Shattuck and Klatt (1976). They presented brief pure-tone acoustic patterns to adult listeners who had to make a similarity judgment. The acoustic patterns were two-component frequency glissandos with a short-term spectral composition similar to the formant transitions in speech.

The results of these experiments did not support the original hypothesis that mirror-image acoustic patterns are intrinsically similar for a listener. Instead, judgements of perceptual similarity for these patterns were based primarily on

1) This research was supported by NIMH research grant MH-24027-04, NIMH Post-doctoral fellowship MH-5823-03 to MEG and a fellowship from the Guggenheim Foundation to DBP.

2) Currently at the Speech Group, Research Laboratory of Electronics, M.I.T. Cambridge, Mass.

the direction of the lower glissando component, the component occurring in the region of the second formant.

We have conducted three experiments that also address the question of whether mirror-image acoustic patterns are intrinsically related. However, we used acoustic patterns that included a steady-state constant frequency (CF) portion as well as a rapid frequency glissando (FM). In addition, we used a perceptual learning paradigm in which listeners had to learn to map four different acoustic patterns into two response categories. We wanted to know which of several mapping arrangements of these patterns would be easiest for listeners to learn.









Stimuli

Three sets of stimuli, with four signals per set, were generated using a complex-tone generating program (Kewley-Port, 1976). Each stimulus component consisted of a 60 msec linear rise or fall (FM) in frequency and a 140 msec constant-frequency (CF) portion. The four signals within a set differed in whether the frequency transition was rising or falling and whether the transition preceded or followed the steady-state portion.

The three stimulus sets differed in the number of component tones, either one, two or three. Frequency values were selected to correspond to values of the first, second, and third formants in the syllables /ba/, /da/, /ab/, and /ad/. For the Single-Tone set, the patterns corresponded to the frequency of the second formant. For the Double-Tone set, component frequencies corresponded to the second and third formants. For the Triple-Tone set, all three formants were represented although the frequency transitions corresponding to the first formant always rose when it preceded the steady-state and fell when it followed the steady-state, in accordance with the formant motions observed in natural speech.

Experimental Procedure and Design

In the perceptual learning task one stimulus from a particular set was presented via headphones on each trial to subjects who responded by pressing one of two response buttons. Correct feedback was provided after each response according to one of three stimulus mapping arrangements. In the Mirror-Image

condition, stimuli with a rising transition preceding the steady-state () or a falling transition following the steady-state () were assigned to one response (R1) whereas stimuli with a falling transition preceding the steady-state () or a rising transition following the steady-state () were assigned to the other response (R2). In the Rise-Fall condition, stimuli with rising transitions either preceding or following the steady-state () were assigned to one response: the two stimuli with falling transitions () were assigned to the other. In the Temporal-Position condition, the stimuli were assigned to responses according to the temporal position of the transitions -- whether the transition preceded () or followed () the steady-state frequency.

In addition to test trials on which responses were collected, study periods were also interspersed to help subjects learn the appropriate stimulus-response mapping. During study periods, several repetitions of each of the test stimuli were presented while feedback was provided.

Results and Discussion

Responses were analyzed in terms of number correct by stimulus. Of the three mapping conditions, the Temporal-Position condition showed the highest performance with 88.4% correct. More importantly, however, the Mirror-Image condition produced more accurate responding, 78% correct, than the Rise-Fall mapping condition with 68.3% correct response.





With respect to stimulus set, the Double-Tone stimuli showed slightly better performance, 81.6%, than the Single-Tone set, with 78.7%. However, the Triple-Tone set which more closely resembled speech showed the poorest performance with only 74.5% correct.

These results have several implications for learning of acoustical patterns resembling speech. First, the easiest stimulus mapping condition to learn was the one based on the temporal position of the transition in the pattern -- a relationship that clearly does not require the subject to analyze out the component frequencies at onset or offset. Subjects can simply use temporal position (i.e., initial vs. final) as the

most salient dimension for learning and ignore all other differences. Second, it is also apparent from our results that when the position of the transition becomes an irrelevant attribute to be ignored by the subject, a stimulus arrangement based on a mirror-image relationship is, in fact, easier to learn than one based on only the direction of frequency change of the transitions. Thus, while mirror-image patterns are not the same perceptually, subjects are nevertheless able to recognize and selectively attend to the criterial properties of the stimuli that define their equivalence. In both of these mapping conditions, subjects must "hear-out" the individual components of the patterns and respond to them selectively.

In Experiment 1 we used tonal patterns so that the stimuli would not be heard as speech. The results would be uninteresting if subjects simply heard these patterns as speech and used phonetic labels to mediate acquisition. By this interpretation, mirror-image patterns would be superior to direction-of-transition mapping because the stimuli within a mirror-image pair evoke the same phonetic label -- i.e. "b" or "d" -- and not because their configural properties are intrinsically related. To study the effects of categorization on perceptual learning we carried out another experiment to determine how well subjects could identify these patterns when explicitly provided with labels that emphasized attention to either the acoustic or phonetic properties of the stimuli.

Experimental Procedure and Design: Experiment 2

In Experiment 2 subjects were required to identify the stimuli into one of four categories provided by the experimenter. In the Acoustic-Label condition, subjects were told that the stimuli were tones consisting of a short interval with constant pitch, followed or preceded by a rapid rise or fall in pitch. The acoustic labels were schematic line drawings of the time course of the frequency change of each stimulus ( ,  ,  , ). In the Phonetic-Label condition, subjects were told that the stimuli were modified tokens of natural speech and were provided with the labels /ba/, /da/, /ab/, and /ad/.

Results: Experiment 2

For the Single- and Double-Tone stimuli, acoustic labels were matched more accurately than phonetic labels. This effect was reversed, however, for the Triple-Tone stimuli, where phonetic labels were more accurate than acoustic labels. The addition of a component in the region of the first formant markedly increased the accuracy of phonetic categorization while decreasing performance with acoustic labels. Note, however, that the low tone for each Triple-Tone stimulus had either an initial rising or final falling transition which did not parallel the direction of the other transitions. The presence of these conflicting frequency glissandos no doubt produced interference in the acoustic labeling condition.

In Experiment 1 acquisition performance was poorer for Triple-Tone stimuli than for Single- and Double-Tone stimuli, a pattern which was replicated here only in the Acoustic Label conditions. Thus, these results support the interpretation that subjects in Experiment 1 were listening primarily in an auditory, rather than phonetic mode and strongly suggest that phonetic-mediation was not responsible for the outcome observed earlier.

The mirror image patterns used here always shared three properties in common: (1) they had the same steady-state frequencies, (2) the frequency transitions at onset and offset had the same short-term spectral composition, and (3) members of a given mirror-image pair had roughly the same average frequency. These three properties are potential factors that could contribute to the salience of mirror-image pairs and the advantage observed in learning. The first factor could not have played a role in the earlier results since all pairs within a stimulus set, whether mirror-image or not, had identical steady-states. However, the other two properties could have been used as reliable discriminative cues by subjects to facilitate learning.

To determine which of these acoustic attributes, if any, was responsible for the advantage observed in learning mirror-image pairs, we repeated Experiment 2 adding two additional stimulus sets, one with average frequencies and the other with the

transitions adjusted to be equal. We also included the original constant steady-state stimuli. Would the mirror-image condition continue to show an advantage in learning under these two new stimulus conditions?

Experimental Procedure and Design: Experiment 3

The experimental procedure was the same as Experiment 1. However, only Double-Tone stimuli were used and the Temporal-Position mapping condition was eliminated. Three stimulus sets were constructed, again with four stimuli per set. Each set contained the initial-rising and final-falling stimuli used earlier. For the stimuli with identical transitions, frequency values for the initial-falling and final-rising stimuli were set lower than in the previous experiment, so that the transitions were identical for all stimuli. These two stimuli were also lowered in frequency for the Average-Frequency set to make all stimuli equal in average frequency.

Results: Experiment 3

The difference in performance between Mirror-Image and Rise-Fall mapping conditions was replicated with the Constant-Steady-State stimuli. The difference between Mirror-Image and Rise-Fall was, however, even greater for the stimuli containing identical transitions. However, for the Average-Frequency stimuli, no significant difference was found between the two mapping conditions. Thus, on the one hand, the advantage of Mirror-Image over Rise-Fall mapping can be made more pronounced by adjusting the frequency transitions to be identical whereas the difference can be attenuated substantially by adjusting all stimuli to have the same average frequency irrespective of the temporal and spectral properties of the patterns.

Conclusions

Mirror-Image acoustic patterns show an advantage in perceptual learning because subjects respond not only to individual components of these patterns but also to properties of the entire pattern in terms of its configural shape. Subjects do not seem to attend selectively to only the gross shape of the spectrum at onset or offset but prefer instead to integrate and deploy salient cues contained in both the transitional and

steady-state portion of the entire patterns. In the case of Mirror-Image patterns, criterial differences between the responses happen also to be correlated with salient and well-defined redundant properties of the patterns such as average pitch which was an irrelevant and uncorrelated dimension when the patterns were arranged in the Rise-Fall mapping condition.

It is apparent from these results with nonspeech signals having properties similar to speech that differences in "mode of processing" can also control perceptual selectivity and influence the perception of individual components of the stimulus pattern as well as the entire pattern itself. This can occur in quite different ways depending on whether the subject's attention is directed to coding the auditory properties or the phonetic qualities of the patterns. These new results on mirror-image patterns have been obtained in a perceptual learning task despite the report that the perceptual similarity of these acoustic patterns cannot be recognized consciously by subjects as shown by earlier experiments.

References

- Klatt, D.H. and S.R. Shattuck (1975): "Perception of brief stimuli that resemble rapid formant transitions," In G. Fant and M.A.A. Tatham (eds.): Auditory Analysis and Perception of Speech. New York: Academic Press, 294-301.
- Kewley-Port, D. (1976): "A complex-tone generating program," Research on Speech Perception Progress Report No. 3 Department of Psychology, Indiana University, Bloomington, Indiana.
- Shattuck, S.R. and D. Klatt (1976): "The perceptual similarity of mirror-image acoustic patterns in speech," Perception and Psychophysics, 20, 470-474.

DUPLEX PERCEPTION AND INTEGRATION OF CUES: EVIDENCE THAT SPEECH IS DIFFERENT FROM NONSPEECH AND SIMILAR TO LANGUAGE

Alvin M. Liberman, Haskins Laboratories, New Haven, Connecticut

The observations relevant to a comparison of speech and nonspeech -- the subject of our symposium -- can be divided, according to one's purposes, into several classes. As my contribution to our discussion, I would call your attention to two. In the one class are those research findings that can enlighten us about speech as a putative mode of perception, different phenomenologically from other aspects of audition. In the other class are those findings that permit us to see one of the possibly unique characteristics of speech as an instance of a correspondingly unique characteristic of language.

Phonetic perception as a mode. Perhaps the most direct way to observe the characteristic differences between perception in the phonetic and auditory modes is to contrive to have the same stimulus patterns perceived, either alternatively or simultaneously, as speech and nonspeech. The first reported example was by Lane and Schneider (1963). Using more or less unnatural synthetic speech that sampled the continuum of voice-onset-time from voiced to voiceless syllable-initial stops, they undertook to obtain discrimination functions under two conditions: in one, they told the subjects they would hear speech; in the other, arbitrary sounds of a complex sort. In the event, the discrimination functions were different: those obtained in the "speech" condition showed the usual peak at the phonetic boundary, while those from the other condition did not.

A more recent and thoroughgoing experiment of this general type has been carried out by Bailey, et al. (1977). Inasmuch as it is to be reported at this symposium, I will say no more about it.

In both of these studies, stimulus patterns were heard as speech or nonspeech, but not at the same time. Let us turn now to those cases in which the two percepts are experienced simultaneously. Such duplex perception was accomplished first by Rand (1974), and applied by him to perception of the transition cue for syllable-initial stops. Into one ear he put just the brief second- and third-formant transitions that are sufficient,

in the appropriate context, to produce the perceived difference in place among [b], [d], and [g]. By themselves, these transitions sound more like a nonspeech 'chirp' than anything else. Into the other ear, he put the remainder of the pattern. Let us call it the 'base'. By itself, the base sounds more or less like a syllable. To some listeners, indeed, it appears to be a stop-vowel syllable but, having a fixed acoustic structure, it can, of course, produce only one stop, not all three. The interesting effect occurs when the transition cue and the base are presented dichotically (and in approximately the right time relationship), for then the listener fuses the two inputs so as to perceive the three coherent stop-vowel syllables he would have perceived had the two inputs been mixed electronically (and presented binaurally), while at the same time perceiving the 'chirp' he would have perceived had the transition cues been presented in isolation. (The chirp is heard in the ear to which the isolated transitions are presented; the syllable is heard in the other ear.) Thus, the same brain perceives the same stimulus input in two phenomenologically different ways, as speech and nonspeech, at the same time. This provides, at the very least, an excellent way to gain an impression of what the difference between phonetic and auditory modes sounds like.

Being interested in the possibility of using Rand's technique for the purpose of further and more nearly precise comparisons of the phonetic and auditory modes, I succeeded in reproducing the phenomenon, but experienced difficulty in getting it to be sufficiently stable to permit the further investigations I had in mind. Recently, however, David Isenberg and I (1978) have produced a duplex percept like that of Rand, but easier to hear, we think, and also, perhaps, more stable. We followed Rand's procedure, but changed the stimulus pattern. In particular, we thought it advisable to make the critical (and isolated) cue be the third-formant transition. The advantage, in that case, might be that the remainder of the pattern -- all of the first and second formants, plus the steady-state portion of the third formant -- would be quite full and speechlike. Accordingly, we chose the contrast [r] vs [l], putting the critical third-formant transition cue into one ear and the remainder (the base) into the other. It was quickly apparent that this arrangement

did make it relatively easy to obtain the duplex percept. Indeed, tests with a number of listeners confirmed that they could "correctly" hear [r] or [l] (depending on which transition cue was presented), while simultaneously hearing the transition cue as a chirp.

Having thus found that the simultaneous fusion and separation of the two parts of the pattern (base and isolated transition) seemed to occur easily and consistently, we undertook further tests. The one I would briefly describe here was designed to assess the effects on the duplex percept of separately varying the intensity of the two stimulus components. We observed, first, that changing the intensity of the transition cue caused changes in the perceived loudness of the chirp, but no such changes in the fused [ra] or [la]; changes in the perceived loudness of the fused [ra] and [la] seemed to occur only as a result of variations in the intensity of the base.

To test this manifestation of duplexity more systematically, we carried out the following experiment. On each trial, we presented to the listener a sequence of dichotic pairs in which the transition component (in one ear) varied between the [r] cue and the [l] cue according to some predetermined order. Also, on each trial we varied the intensity of (1) the transition cue, or (2) the base, or (3) neither. The listener's task was to tell the order of [ra] and [la] syllables he heard, and, in addition, which loudnesses, if any, had changed. The results indicated that our listeners were, in fact, fusing the dichotic inputs to hear the 'correct' syllable, while at the same time dissociating the loudnesses by assigning the intensity of the transition cue to the chirp and the intensity of the base to the fused speech percept. We find this phenomenon interesting in its own right, but also as a basis for further investigation into the properties of the two components -- speech and nonspeech -- of the duplex percept. Consider, in that connection, an earlier study by Mattingly et al (1971) that compared the discrimination of a transition cue when, in the one case, it was presented in isolation and perceived as a chirp, and when, in the other, it was in its proper place in the speech pattern and cued a phonetic distinction. That study found differences in the discrimination pattern under the two conditions, but the interpretation of that finding

was subject to the reservation that the transition cue was, after all, in different contexts. By taking advantage of the duplex percept, we can, perhaps, obtain results that will avoid the need for that reservation and thus speak more straightforwardly to the difference between speech and nonspeech.

The integration of cues in speech and syntax. One of the most general characteristics of speech is that the information appropriate to a phonetic segment is typically contained in a numerous variety of cues; moreover, these are widely distributed through the signal and sometimes overlapped with cues for other phones. This is so because of the nature of articulation and coarticulation (Cooper, 1963; Fant, 1973): the various components of an articulatory gesture, distributed as they are in time, spread their acoustic consequences through the signal. Thus, the closing and opening gestures appropriate to an intervocalic stop affect the duration of the preceding syllable and also its offset, the occurrence and duration of an intervocalic silence, and the temporal and spectral characteristics of the onset of the following syllable. Conversely, and as a result of coarticulation, information about successive segments is often collapsed into a single acoustic segment and conveyed simultaneously, as in the case of most consonant-vowel syllables. Yet the speech processor somehow sorts the cues, as it were, assigning each to the appropriate part of the perceived phonetic structure. More to the point of our present purpose, it "integrates" into a unitary percept all the cues for a particular phone, no matter how various and widely distributed the cues may be (Lieberman and Studdert-Kennedy, 1977; Repp et al, 1978; Bailey and Summerfield, 1978; and Dorman et al, 1978). It is difficult to see how this can be accomplished by ordinary auditory mechanisms, so we assume phonetic processes specialized for the purpose.

Consider, for example, the above-mentioned experiment by Repp et al. It dealt with perception of the utterance: "Did you see the gray (great) ship (chip)?" The variables of interest were (1) the nature of the next-to-last word, which was biased either toward gray or great; (2) the duration of the silent interval between gray (great) and ship (chip), and (3) the duration of the fricative noise in ship (chip).

Let us now look first at the "forward" action of an earlier-

occurring cue on a later-occurring one: given a perceptual boundary between ship and chip that varied according to the duration of the fricative noise and also the duration of the preceding silent interval, there was a further variation that depended, other things equal, on whether the preceding word was biased toward gray or great. Now consider an effect in the opposite direction -- the "backward" action of a later-occurring cue on the perception of an earlier-occurring one. This was exemplified by the finding that the listener perceived gray or great depending, all else equal, on the duration of the fricative noise in ship; with other cues properly set, the listeners perceived 'gray' when the duration of the fricative noise (in the next syllable) was relatively short, but great when it was relatively long. Thus, the perception of gray could be changed to great by adding fricative noise in the syllable that followed the target word.

Apparently, the listeners in that experiment integrated into a unitary phonetic percept a variety of acoustic cues that stretched over at least two syllables and overlapped completely with cues relevant to other phones. But how does the listener do this? More specifically, how does he know when to stop integrating? Looking at the variety of cases of this type, we conclude that the integration period is marked neither by a temporal criterion (integrate every x msec), nor by an acoustic one (integrate every time a particular kind of sound is heard). Rather, the integration seems to occur over any stretch of the signal that contains the acoustic consequences of just those articulatory maneuvers that are the peripheral reflections of the speaker's intent to produce a particular phonetic segment. We must wonder, then, how the listener delimits the proper span over which to integrate, in what form he holds the pre-integrated cues, and what he does while waiting.

Consider now, though briefly, how analogous this is to what happens in the decoding of syntax. Surely, the meaning of a syntactic structure (e.g., a sentence) cannot be had except as the listener takes account of the words the structure comprises. As in the phonetic case, the size of this structure is not defined by a temporal criterion, nor by an acoustic one. Rather, it appears to be any number of words that are relevant to the

syntactic structure, and that depends, in turn, on the nature of the message the speaker means to convey. Here, too, then, we must wonder how the listener knows when the structure is complete, in which form he holds the words pending completion, and what he does while waiting.

References

- Bailey, P.J., Q. Summerfield, and M. Dorman (1977): "On the identification of sine-wave analogues of certain speech sounds", Haskins Laboratories Status Report on Speech Research, SR-51/52, 1-25.
- Bailey, P.J. and Q. Summerfield (1978): "Some observations on the perception of [s] + stop clusters", Haskins Laboratories Status Report on Speech Research, SR-53, Vol. 2, 25-60.
- Cooper, F.S. (1963): "Speech from stored data", 1963 IEEE International Convention Record, Part 7, p. 139.
- Dorman, M., L. Raphael, and A.M. Liberman (1978): "Some experiments on the sound of silence in phonetic perception", (submitted for publication).
- Fant, G. (1973): "Descriptive analysis of the acoustic aspects of speech", Speech Sounds and Features, Ch. 2, 25-6. (Article based on a paper by Fant presented at a Wenner-Gren Foundation Research Symposium held at Burg Wartenstein, Austria, 1960, which appeared originally in Logos, 5, 3-17 (1962).)
- Isenberg, D. and A.M. Liberman (1978): "Speech and nonspeech percepts from the same sound", JASA 64, Suppl. No. 1, J20.
- Lane, H.L. and B.A. Schneider (1963): "Discriminative control of concurrent responses by the intensity, duration and relative onset time of auditory stimuli", unpublished report, Behavior Analysis Laboratory, University of Michigan.
- Liberman, A.M. and M. Studdert-Kennedy (1977): "Phonetic perception", in Handbook of Sensory Physiology, Vol. VIII, "Perception." ed. by R. Held, H. Leibowitz, and H.L. Teuber; Heidelberg: Springer-Verlag, Inc.
- Mattingly, I.G., A.M. Liberman, A.K. Syrdal, and T. Halwes (1971): "Discrimination in speech and nonspeech modes", Cogn. Psych. 2, 131-157.
- Rand, T.C. (1974): "Dichotic release from masking for speech", JASA 55, 678-680.
- Repp, B.H., A.M. Liberman, T. Eccardt, and D. Pesetsky (1978): "Perceptual integration of temporal cues for stop, fricative, and affricate manner". J. Exp. Psych.: Human Perception and Performance (in press).

ISSUES IN SPEECH PERCEPTION

Dominic W. Massaro, Department of Psychology
University of Wisconsin, Madison, Wisconsin, 53706, USA

My goal in the present paper is to address what I believe to be some important issues in the study of perception of speech and nonspeech sounds. The issues are discussed in the framework of binary contrasts. The binary framework was deemed appropriate because of both linguistic precedent and limited psychological capacity. Some hierarchical organization of the issues is probably optimal but I have been reluctant to provide one; the reader can sort, add to, delete, and order the issues as she or he chooses.

Templates versus features

Speech sounds may be gestalt units that cannot be further analyzed or reduced in terms of other attributes. If speech consisted of a sequence of indivisible sounds, then speech analysis would be limited to some variation of a template matching scheme. For successful analysis, an additional template would be needed for every unique speech sound. Although this possibility may be linguistically and psychologically correct, it leaves the student very little to do beyond a general recording and tabulation.

Not only does the template matching scheme leave time on the student's hands, it is not very appealing to those of us who wish to impose simplicity and order upon Mother Nature (or Mother Tongue). Luckily, Jakobson and his colleagues of the Prague school successfully argued that phoneme units could in fact be further analyzed in terms of distinctive features that represent similarities and differences with respect to other phonemes. Given this theoretical perspective, it follows naturally that all of the phonemes of a language can be characterized in terms of a set of distinctive features. Feature analysis is appealing because it allows the units to be subjected to a more abstract classification.

Feature analysis is also preferred over template matching in the study of perception. Template matching schemes would not illuminate any perceived similarities or differences among speech sounds. The applicability of feature analysis proves useful in understanding the findings that two sounds are perceived as similar to one another or are in fact confused with one another to the extent they share the same features. Independent evidence for

features comes from well-known neurophysiological findings that individual cells in the cortex respond selectively to a class of stimuli that share a particular property, such as the direction of the frequency change in a sound. Feature analysis is a worthwhile enterprise as long as we sometimes remind ourselves that it must stop somewhere. When a set of descriptors is no longer analyzable we are left with miniature templates.

Binary versus continuous features

Although it is not unreasonable to describe a speech sound in terms of the degree to which a feature is present in the sound, Jakobson made the important assumption that distinctive features¹ were binary in that each feature is either present or absent in all-or-none fashion. Jakobson argued that "the dichotomous scale is superimposed by language upon the sound matter." The idea of binary features is appealing in terms of parsimony but most importantly in terms of ease of classification. The integration of binary information from two or more feature dimensions requires only logical conjunction of pluses and minuses. The elegance of binary classification is probably responsible for what might sometimes be viewed as an excessive observance of the principle.

In terms of speech perception, it seems more reasonable to assume that the listener has information about the degree to which each feature is present in the speech sound. This assumption of continuous rather than all-or-none featural information contrasts with the traditional view of binary features in linguistic theory. More recently, Chomsky and Halle and Ladefoged have allowed a multi-valued representation of featural information at the perceptual level. In our model, each feature is evaluated in terms of a fuzzy predicate that specifies the degree to which it is true that the sound has a particular feature. Given the fuzzy information passed on by feature evaluation, it is apparent that the integration of this information across several features is more complex than in traditional all-or-none classificatory schemes. Much of our work has supported the idea that features are combined in

(1) The reader should be reminded that the issue of binary versus continuous features is independent of other issues such as phonetic versus acoustic features. Accordingly, even though some examples are drawn from linguistic analyses, the use of features is intended to be general and not limited to one level of analysis.

terms of a multiplicative rule. This combinatorial process is extremely simple but has the nice consequence that the less ambiguous features carry more weight.

Phonetic versus acoustic features

It is readily transparent that the concept of phonetic features has advanced the study of the linguistic classification of speech sounds. Students of speech perception must further inquire, however, whether speech perception is mediated by phonetic and/or acoustic features. The seminal work at Haskins Laboratories using synthetic speech evolved around the assumption that phonetic features were perceptually real. Many experiments were carried out to determine which acoustic properties of speech sounds were responsible for the perceived presence or absence of phonetic features. Given our analysis in the discussion of templates versus features it follows that the acoustic properties of speech sounds could be evaluated in terms of templates or features. If you agree that feature analysis is more desirable, then the speech perception theorist must be concerned with the analysis of speech sounds in terms of acoustic, not just phonetic, features.

Single factor versus multifactor experiments

In most experiments, speech sounds are varied along a single relevant dimension and observers are asked to perceive a given contrast between two sounds. For example, in the study of the acoustic features for a voicing contrast, all acoustic properties relevant to the contrast are made relatively natural except one, such as voice onset time, and this property dimension is varied through a continuum of values. Very few experiments independently vary more than one property within a particular experiment. The few exceptions in the early literature essentially reduced the data analysis to single-property experiments. In our work we utilize factorial designs and functional measurement techniques to study how acoustic features are evaluated and integrated together. With this procedure, two or more acoustic dimensions are independently varied so that all combinations of the values of one property are paired with all combinations of the values of another property. This design allows a direct assessment of how the acoustic features are evaluated and integrated together in speech perception.

Independent versus dependent features

This issue centers around whether the value for a given feature is modified by the value of another feature. Some support for featural independence was provided by studies demonstrating that separate sets of acoustic properties were relevant for perception of different contrasts. However, this result does not necessarily rule out the possibility that the perception of one contrast is dependent on the perception of another. Nonindependence has been proposed to account for the observed shifts in a voicing-contrast boundary as a function of a contrast in terms of place of articulation. However, these boundary shifts may occur even if each of the features makes independent contributions to the analyses. The observed interaction may result from the manner in which the independent featural information is integrated together. A quantitative model based on this idea has been successful in providing a quantitative account of boundary shifts and, therefore, the shifts do not imply nonindependence of feature evaluation.

Phoneme versus syllable units

Speech sounds of phoneme size have proven to be valuable in linguistic analysis. For the student of speech perception, however, it is important to ask what sound units are perceptually real. Although it is not easy to determine the sound units that are functional in speech perception, the question can be addressed simultaneously with the study of acoustic features in speech perception.

In our model, features are evaluated and matched to those features which define units in long-term memory. A unit is represented in long-term memory by a prototype which consists of a list of acoustic features. We assume that perceptual recognition of speech is mediated by vowel, consonant-vowel, or vowel-consonant syllable units in long-term memory. This assumption contrasts with the more commonly accepted notion of phonetic or phonemic prototypes in which phonetic or phonemic decisions mediate speech perception. Although it is only natural to say that a particular acoustic property cues voicing, the perception of the phonetic feature of voicing does not mediate syllable recognition in our model. Experiments that have evaluated the acoustic properties that are responsible for phonetic contrasts ask listeners to distin-

guish among speech segments of, at least, syllable length. These experiments do not necessarily mean that speech perception of the syllables was mediated by the phonetic contrasts defined by the experimenter.

In addition to the problem of the lack of acoustic invariance for some consonants, phoneme units cannot easily account for the finding that the vowel sometimes provides direct acoustic information about the consonant portion of a syllable. Vowel duration has a large effect on the voicing contrast of a vowel-consonant syllable in word-final position. Experimental and theoretical work in our laboratory supports the idea that acoustic features of the vowel portion and consonant portion are perceived independently, integrated together, and evaluated against syllable units in memory.

Stimulus versus process descriptions

Researchers are converging on the belief that there exists a plethora of potential acoustic features in speech perception. In contrast to the relatively small number of linguistic distinctive features, the potential candidates for acoustic features seem endless. Faced with this army of potential features, what might be the most valuable tack to take? Rather than attempting to define and catalog the large family of features, it might be more worthwhile to design prototypical experiments to assess how a small number of acoustic features are evaluated and integrated together in speech perception. The goal would be to develop a testable description of the process of speech perception rather than a complete stimulus description of all acoustic features. Needless to say, good judgment on the part of the speech researchers will allow a gradual accumulation of a stimulus description in their quest for understanding speech perception processes.

Acoustic versus contextual determinants

Speech perception research has been characterized by the study of speech perception as a function of acoustic changes in speech sounds. The researchers have not denied that other sources of information may also be exploited in perceiving natural speech. Not long after the investigator begins to understand how acoustic features are evaluated and integrated together in speech perception, it becomes necessary to assess how the processes work when

contextual influences are also available. As an example, feature evaluation and integration could be studied as a function of both acoustic changes in the speech signal and contextual constraints in terms of how likely a given sound may occur in a given context. A quantitative description of analogous experiments in reading supports the idea that contextual constraints simply provide an independent source of information exactly analogous to what would be provided by an additional feature.

Speech perception versus speech recognition

Upon reflection, it is apparent that speech recognition does not mirror speech perception. I recognize (and classify) two sounds as the same without necessarily perceiving them as identical. I believe that the idea of perceptual constancies has misled researchers in not only areas of visual perception but also in speech. The receding object is recognized as the same object even though the retinal input undergoes drastic changes. But the perception of the object also changes as is easily demonstrated by a little perceptual scrutiny. Following in the behavioristic tradition, researchers usually ask listeners to identify or classify sounds and take performance as an index of perception. Are we asking observers to make the stimulus error as the early introspectionists would claim or are there experimental tasks and performance measures that provide good indices of speech perception? This issue may help illuminate the general area of categorical perception by asking to what extent categorical perception is not categorical perception but simply categorical recognition.

To more directly tap perception, experimenters might employ continuous rather than discrete response alternatives. A discrete judgment may not be sensitive to the continuous changes in perception produced by continuous changes in an acoustic property of the speech sound. As an example, small increases in voice onset time for a velar stop might be perceived as making the sound more like /ki/. However, if the sound is still perceived as more like /gi/ than /ki/, the listener may always respond with /gi/. If the listener's judgments are consistent, the different sounds would be responded to equivalently even though they are perceived as different. By asking the observer to make a judgment on a continuum between the discrete alternatives, the responses may

more directly mirror perception. We have obtained orderly data from observers marking off a line in order to place the percept somewhere between discrete alternatives.

Speech perception versus speech understanding

It is easy to forget that speech perception does not necessarily entail speech understanding and that accurate understanding does not demand accurate speech perception. Consider a lexical decision task in which a listener indicates whether each test is a word or a nonword. The nonwords, such as "prust" and "mantiness", are perceived correctly and could be repeated even though no understanding takes place. I don't think that it would be profitable to argue that nonwords are not perceived. Our last noisy party reminds us that a significant amount of speech understanding can occur without perfectly accurate speech perception. In many highly constrained sentence contexts, the listener understands exactly some of the message before he perceives it. In fact, a few recent studies have provided some support for the idea that understanding can actually modify perception. A more convincing demonstration is how the perceived clarity of the words of a song is enhanced when the listener simultaneously reads them. In any case, it is necessary to distinguish between the case in which the listener resolves a piano sound sufficiently to distinguish it from adjacent sounds on the musical scale and the case in which the sound is also identified as middle C.

In our model, perception and understanding occur at two different stages of information processing. The primary recognition process evaluates and integrates acoustic features and outputs a perceptual experience of a speech sound. The secondary recognition process operates on the perceptual information to impose meaning and, therefore, a relatively abstract encoding. Although these are highly analogous processes, they utilize different categories of information in long-term memory and may be influenced by different properties of higher-order contextual constraints.

Speech versus nonspeech

It seems appropriate to close with this issue (or nonissue) since it is the topic of this symposium. Although speech represents language and nonspeech does not, it is important to know to what extent perception of speech is analogous to perception of

nonspeech. Does nonspeech perception derive from an evaluation and integration of acoustic features defined with respect to segments of sound? Remarkable parallels between speech and nonspeech have been reported in recent years. Rather than concluding that serious investigators should return to psychophysical studies of nonspeech in order to understand basic auditory processes, it seems more productive to assume that speech offers so much more for experimental study and that the most direct route to an understanding of auditory perception is to be found in the study of speech perception.²

References

- Derr, M.A. and D.W. Massaro (1978): "The contribution of vowel duration, F₀ contour, and frication duration as cues to the /juz/-/jus/ distinction." WHIPP Report #8.
- Massaro, D.W. (1978): "Letter information and orthographic context in word perception." Technical Report No. 453.
- Massaro, D.W. (1975): (Ed.) Understanding language: An information processing analysis of speech perception, reading, and psycholinguistics. New York: Academic Press.
- Massaro, D.W. and M.M. Cohen (1976): "The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction", JASA, 60, 704-717.
- Massaro, D.W. and M.M. Cohen (1977): "The contribution of voice-onset time and fundamental frequency as cues to the /zi/-/si/ distinction", Perc. Psych. 22, 373-382.
- Massaro, D.W. and G.C. Oden (1978): "Evaluation and integration of acoustic features in speech perception", WHIPP Report #9.
- Oden, G.C. (1978): "Integration of place and voicing information in the identification of synthetic stop consonants," JPh, in press.
- Oden, G.C. and D.W. Massaro (1978): "Integration of featural information in speech perception", Psych. Rev. 85, 172-191.

(2) The research reported in this paper was carried out with the collaboration of Michael M. Cohen, Marcia A. Derr, and Gregg C. Oden and was supported in part by National Institute of Mental Health Grant MH 19399 and in part by the Wisconsin Alumni Research Foundation.

WHAT TELLS US THAT SPEECH IS SPEECH?

Quentin Summerfield, MRC Institute of Hearing Research, University Medical School, Nottingham, UK, and Peter J. Bailey, Department of Psychology, University of York, York, UK.

Acoustic analysis and perceptual experimentation have suggested that speech sounds are special and distinct from other sounds. First, no obvious one-to-one isomorphism exists between acoustic and phonetic segments, and the latter have been said to be encoded in the former (e.g., Liberman et al., 1967). Secondly, phonetic perception is apparently not fully rationalised by known psycho-acoustic properties of the auditory system. One illustration of this is provided by the experiments described by Dr Dorman (this symposium; see also Bailey, Summerfield and Dorman, 1977), in which the perception of sinewave analogues of speech is shown to depend upon listeners' interpretation of the signals as speechlike or non-speechlike. There is thus some support for the argument that speech perception entails a special decoding process (Liberman & Studdert-Kennedy, 1977). In what follows we shall explore two related questions: what is the nature of the information which might activate such a process, and to what extent should a specification of this information constrain a formulation of the process?

One simple hypothesis could be that speechlikeness is marked by acoustical attributes which, if detected in an initial stage of auditory analysis, direct the signal to a subsequent stage of special phonetic processing. Such attributes would have to be properties of all utterances, and, of necessity, would have to be unencoded, unlike the contextually mutable segments whose decoding they would trigger. Possible candidates have been considered to be rapid spectral changes (Haggard, 1971) and the onset of periodic excitation (Allen & Haggard, 1977), but their role as 'trigger features' has not been empirically demonstrable. Furthermore, even if elaborated by a variable criterion for the acceptability of a trigger feature, this hypothesis cannot account for the perceptual duality of sinewave analogues of CV syllables. Here the putative trigger features would have to be intrinsic to the information specifying phonetic identity, and so in failing to meet the criterion of invariance would exceed the categorising capabilities of purely auditory analysis. Paradoxically, to detect such trigger features

successfully, the putatively auditory processor would require the properties of a special phonetic decoder.

These considerations call for a re-appraisal of the model in which signals are routed to one type of processor or another on the basis of prior detection of simple acoustical attributes. They suggest that an alternative solution to the problem of distinguishing speech from non-speech sounds could be that phonetic and generalised auditory processing are accorded in parallel to all acoustic inputs. Phenomenal perception would correspond to whichever process achieved a satisfactory analysis. In proposing such a solution, Liberman, Mattingly and Turvey (1972) suggested that a sound is recognised as speech if a phonetic processor succeeds in extracting phonetic features. Thus the acoustic specification of signals as speechlike is conceived as being isomorphic with the acoustic specification of the cues to phonetic elements, and a characterisation of the former would follow inevitably from a characterisation of the latter. We have already noted that speech is considered an intractable perceptual problem, as a result of the non-invariant relationship between acoustic cues and phonetic elements. This contrasts with the more straightforward relationship existing between phonetic elements and articulatory dynamics, which has led to the suggestion that acoustic cues are interpreted with respect to an internalised knowledge of vocal tract behaviour (Stevens & House, 1972; Liberman et al., 1967). A perceptual model of this kind would seem to involve at least two stages: in the first, a sequence of acoustic elements must be segregated and detected; in the second, these elements must be interpreted, presumably to reconstruct the information encoded in the sequential properties of the signal. Knowledge of vocal tract behaviour may assist the first stage, but it governs the second stage. While we have no doubt that speech perception is inextricably tied to the origin of the signal in a vocal tract, we wonder whether a process of fractionation followed by reintegration would best capture the information endowed in the signal by the continuous articulatory flow of a dynamic vocal tract (see also, Bailey & Summerfield, 1978).

It has been the general conclusion of students of perception that distal events and proximal stimulation relate equivocally, and the traditional response to this problem has been to assert that perception is a constructive process mediated by abstract internal

knowledge (see, for instance, Neisser, 1967). This view of perception is currently coming under increasing scrutiny, urging a re-examination of the peculiarities of phonetic perception. Theoretical appraisal (e.g., Turvey, 1977; Shaw & Bransford, 1977) and empirical analysis (e.g., Lee, 1974; Blumstein & Stevens, 1978) suggest that distal events may have a more veridical, if complex, representation in perceptual data than has generally been supposed. Thus it may be profitable to explore the notion that phonetic percepts are not constructed from discrete acoustic elements by the mediation of articulatory knowledge, but rather that they are specified in acoustic dynamics structured by a speech-specific organisation of the vocal apparatus (e.g., Krmpotic, 1959; Fowler et al., in press). The acoustic signal must remain the focus of our concern, given that an unequivocal reconstruction of articulatory dynamics from the acoustic signal is not possible (e.g., Atal et al., 1978).

Implicit in the foregoing is the assumption that information for speechlikeness can be specified at a single level of analysis, for which the most promising popular candidate has been the level of phonetic processing. This is a necessary view, given that listeners can describe as speechlike even highly schematic analogues of speech sounds, provided they permit a phonetic interpretation. However, the notion that speechlikeness is specified only in the information for phonetic elements is insufficient to account for the certainty and immediacy with which naive listeners can identify utterances in an unfamiliar language as human speech. In recognising as speech snatches of foreign languages heard, for instance, when tuning a radio receiver, we are presumably attending to information of a different kind from that which specifies a sinewave analogue of a CV syllable as speechlike. A particular suggestion by Stevens and House (1972) is that natural speech sounds are characterised by 'certain dynamic or time-varying properties, among which are syllabic intensity fluctuations such as are associated with one of the most fundamental attributes of speech - the vowel-consonant dichotomy' (p. 13). Recent reformulations of the processes underlying speech production (e.g., Fowler et al., in press) provide a means of rationalising the multiplicity of information in a speech signal that specifies it as such. In this view, the speaker progressively organises his articulatory musculature such

that moment-to-moment control need only be exercised over the minimum number of muscle groups during the act of speech production. It is suggested that speech is the concomitant of a set of functionally nested constraints upon the organisation of the vocal apparatus as a whole, so that short-term events like consonantal articulations are nested within longer-term events like the reconfiguration of the vocal tract for successive vowels; these are themselves nested within events of even longer life-spans, such as the speech-specific respiratory synergism (e.g., Lenneberg, 1967). All of these articulatory events are characteristic of speech production, and all endow the speech signal with distinctive dynamic properties to which listeners may be sensitive.

This conceptualisation of speech production, and the type of perceptual attunement it implies, are consistent with a broader view of the development of sensitivity to sound in general. In the natural world, sounds result from the participation of three-dimensional structures in events that occur over time. It is held that the evolution in organisms of sensitivity to vibration in the media that surround them progressed as a developing facility in identifying not just vibration or sound per se, but ecologically relevant events whose concomitants are sounds (see, for instance, Masterson & Diamond, 1973). To a greater or lesser degree, a natural sound is specific to (though not necessarily completely descriptive of) both its particular source, and the particular event in which the source is participating.

Following Turvey and Prindle (1978), therefore, we suggest that the distinction typically made in the laboratory between perception of natural (or even synthetic) speech sounds, and perception of non-natural waveforms like isolated pure or complex tones, should be recast as the distinction between the perception of events and the perception of non-events. In terms of this categorisation, speech perception is a particular instance of event perception, and a general description of the auditory perception of natural events should throw light on the specific problem of perceiving articulatory events. A tentative description could be that the perception of events depends upon the registration of the coherence of information specific to a source and information specific to the transformation wrought upon that source. (See Shaw & Pittenger, 1977.) Thus a preliminary answer to the question of

what is a speech sound could be this: a pattern of sound may be perceived as speech if it cospecifies its source as a human vocal tract participating in a physiologically and phonologically permissible act of articulation. The registration of coherence is analogous to perceiving the solutions to a set of simultaneous equations: the equations provide structure and coherence for the solutions, but no one solution necessarily mediates the attainment of any other. What we understand by coherence may be illustrated further with a visual analogy. When a man runs, he structures light in such a way that both his identity as a man and his act of running are specified optically. When we perceive him running, we detect the coherence of these conjoint specifications; we do not first perceive the actor in order that we may interpret the elements of his act. (For a particularly succinct demonstration of the registration of coherence in the perception of such events, see Johansson, 1974.)

It will be apparent that we lack a formal means of characterising the coherence in speech sounds. Nevertheless, the notion provides us with an appealing informal account of the perceptual strategies adopted by listeners in the experiments on sinewave analogues of speech. When sinewaves were heard as non-speech sounds, we suppose that listeners attended to the elements in the acoustic array but not to their potential organisation. In hearing them as speechlike, on the other hand, they attended both to the acoustic elements and to their organisation, which together specify, albeit in a highly reduced form, a vocal tract undergoing a phonologically permissible act of articulation. Those familiar with R.C. James' photograph, reproduced in Lindsay and Norman (1972, p. 8) will recognise that the foregoing analogously describes the initial perception of the picture as a random array of dark and light areas, and the subsequent perception of a Dalmatian dog walking in dappled sunlight. Both hearing sinewaves as speechlike and seeing the dog are compelling perceptions. It may be that the search for coherence in stimulus information is a general goal of perceptual systems, guided and rewarded by the attainment of clarity (Woodworth, 1947; Gibson, 1969). We note that when listeners began to hear sinewaves as speechlike, their identification functions became more consistent and more categorical.

In summary, we are suggesting that the achievement of speech articulation is to present the information for speech perception unequivocally in the surrounding media. The acoustic signal is clearly the most important vehicle for speech, but we acknowledge also the perceptual importance of the speech-specific optical concomitants of articulatory events (e.g., Miller & Nicely, 1955; see Erber, 1975, for a review). Progress beyond the phenomenological interest of demonstrations such as the perceptual duality of sine-wave analogues of CV syllables requires the development of a vocabulary with which to describe how articulatory events structure sound and light in perceptually accessible ways. The mathematics of this description will be complex. Nevertheless, we are encouraged that optical invariants supporting the visual perception of aspects of one human activity, locomotion, have been formally described (Lee, 1974; Cutting et al., 1978). The rebirth of articulatory synthesis for perceptual experimentation (Mermelstein & Rubin, 1978; cf., Haggard, in press) is one precursor of the attainment of a similar specification of the optical and acoustical invariants supporting the perception of speech articulation: that is, to specify what it is that tells us that speech is speech.

References

- Allen, J. and M.P. Haggard (1977): "Perception of voicing and place features in whispered speech: a dichotic choice analysis", Perc. Psych. 21, 315-322.
- Atal, B.S., J.J. Chang, M.V. Mathews, and J.W. Turkey (1978): "Inversion of articulatory to acoustic transformation in the vocal tract by a computer sorting technique", JASA 63, 1535-1555.
- Bailey, P.J. and A.Q. Summerfield (1978): "Some observations on the perception of [s]+stop clusters", Haskins Laboratories Status Report on Speech Research SR53 (2), 25-60.
- Bailey, P.J., A.Q. Summerfield, and M.F. Dorman (1977): "On the identification of sine-wave analogues of certain speech sounds", Haskins Laboratories Status Report on Speech Research SR51-52, 1-25.
- Blumstein, S.E. and K.N. Stevens (1977): "Acoustic invariance for place of articulation in stops and nasals across syllable contexts", JASA 62, S26(A).
- Cutting, J.E., D.R. Proffitt, and L.T. Kozlowski (1978): "A bio-mechanical invariant for gait perception", J. Exp. Psych: HPP 4, 357-372.
- Erber, N.P. (1975): "Audio-visual perception of speech", JSHD 40, 481-492.
- Fowler, C.A., P. Rubin, R.E. Remez, and M.T. Turvey (in press): "Implication for speech production of a general theory of action", in Language production, B. Butterworth (ed.), New York: Academic Press.

- Gibson, E.J. (1969): Principles of perceptual learning and development, New York: Appleton.
- Haggard, M.P. (1971): "Encoding and the REA for speech signals", Quart. J. Exp. Psych. 23, 34-45.
- Haggard, M.P. (in press): "Experience and perspectives in articulatory synthesis", in Frontiers of speech communication research, B. Lindblom and S. Ohman (eds.), London: Academic Press.
- Johansson, G. (1974): "Projective transformations as determining visual space perception", in Essays in honor of J.J. Gibson, R.B. MacLeod and H.L. Pick (eds.), 117-138, Ithaca: Cornell University Press.
- Krmpotic, J. (1959): "Donnés anatomiques et histologiques relatives aux effecteurs laryngo-pharyngo-buccaux", Revue Lar. Otol. Rhinol. 80, 829-848.
- Lee, D.N. (1974): "Visual information during locomotion", in Essays in honor of J.J. Gibson, R.B. MacLeod and H.L. Pick (eds.), 250-267, Ithaca: Cornell University Press.
- Lenneberg, E.H. (1967): Biological foundations of language, New York: Wiley.
- Lieberman, A.M., F.S. Cooper, D.P. Shankweiler, and M. Studdert-Kennedy (1967): "Perception of the speech code", Psych. Rev. 74, 431-461.
- Lieberman, A.M., I.G. Mattingly, and M.T. Turvey (1972): "Language codes and memory codes", in Coding processes in human memory, A.W. Melton and E. Martin (eds.), 307-334, New York: Winston.
- Lieberman, A.M. and M. Studdert-Kennedy (1977): "Phonetic perception", Haskins Laboratories Status Report on Speech Research SR50, 21-60. (To appear in Handbook of Sensory Physiology, Vol. VIII, "Perception", R. Held, H. Leibowitz, and H-L. Teuber (eds.), Heidelberg: Springer Verlag.
- Lindsay, P.H. and D.A. Norman (1972): Human information processing: An introduction to psychology, New York: Academic Press.
- Masterson, B. and I.T. Diamond (1973): "Hearing: central neural mechanisms", in Handbook of perception, Vol. III, Biology of Perceptual Systems, E.C. Carterette and M.P. Friedman (eds.), 408-448, New York: Academic Press.
- Mermelstein, P. and P. Rubin (1978): "Articulatory synthesis - a tool for the perceptual evaluation of articulatory gestures", Haskins Laboratories Status Report on Speech Research SR53 (1), 1-11.
- Miller, G.A. and P.E. Nicely (1955): "An analysis of perceptual confusions among some English consonants", JASA 27, 338-352.
- Neisser, U. (1967): Cognitive psychology, New York: Appleton.
- Shaw, R. and J. Pittenger (1977): "Perceiving the face of change in changing faces: implications for a theory of object perception", in Perceiving, acting and knowing, R. Shaw and J. Bransford (eds.), 103-132, Hillsdale, N.J.: Erlbaum.
- Shaw, R. and J. Bransford (1977): "Psychological approaches to the problem of knowledge", in Perceiving, acting and knowing, R. Shaw and J. Bransford (eds.), 1-39, Hillsdale, N.J.: Erlbaum.

- Stevens, K.N. and A.S. House (1972): "Speech perception", in Foundations of modern auditory theory, Vol. II, J.V. Tobias (ed.), 1-62, New York: Academic Press.
- Turvey, M.T. (1977): "Contrasting orientations to the theory of visual information processing", Psych. Rev. 84, 67-88.
- Turvey, M.T. and S.S. Prindle (1978): "Modes of perceiving: abstracts, comments and notes", in Modes of perceiving and processing information, H.L. Pick and E. Saltzman (eds.), 205-224, Hillsdale, N.J.: Erlbaum.
- Woodworth, R.S. (1947): "Reinforcement of perception", AJPs 60, 119-124.

THE PERCEPTION OF CHINESE SPEECH SOUNDS IN MASKING NOISE
AND FREQUENCY DISTORTION

Tze-Wei Pao and Yung-Tzue Wei, Acoustical Institute of Nanking
University

Intelligibility tests of Chinese speech sounds were run under five masking conditions, namely white noise, pink noise, speech noise, meaningful speech interference, and reverberation masking in an auditorium, as well as in a quiet studio. To simulate the actual communication circumstances, the noise was introduced at input and output ends, respectively. The signal to noise ratios were 5, 0, -5, -10 dB with a fixed speech level about 80 dB at 1m from the loudspeaker. In addition, the speech and noise were processed with high pass, low pass, or band pass filtering except in the reverberation condition. A set of simplified but rather sensitive word lists were used, which were based on varying the initial consonants (initial consonants are more sensitive to masking than are final consonants). The effects of masking and frequency distortion on the perception of individual Chinese speech sounds will be presented in this report.