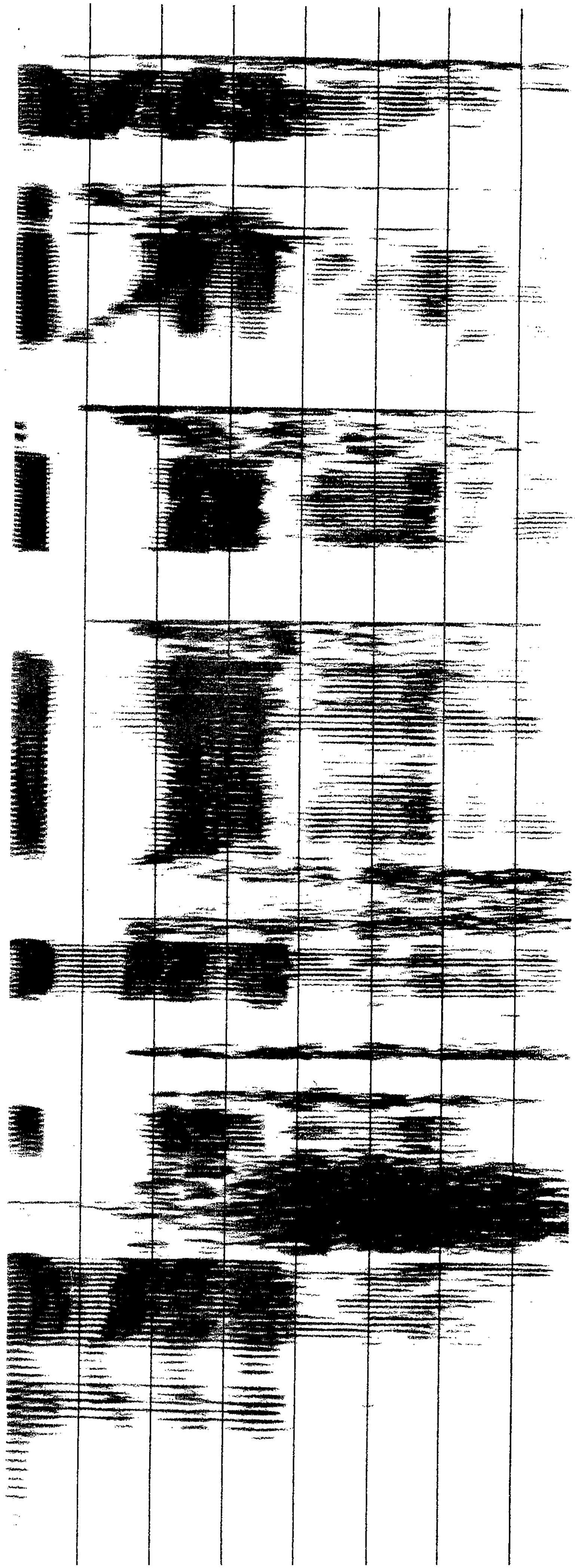
WPP

UCLA WORKING PAPERS IN PHONETICS

APRIL 1983

57



UCLA Working Papers in Phonetics 57

April 1983

The UCLA Phonetics Laboratory Group.

Norma Antoñanzas
Sarah Dart
Anthony Davey*
Karen Emmorey
Vicki Fromkin
Ellen Jackson
Hector Javkin
Bruce Hayes
Nonie Holz
Marie Huffman
Pat Keating
Paul Kirk
Jody Kreiman
Peter Ladefoged
Jenny Ladefoged
Thomas Lee
Mona Lindau
Wendy Linker
Shannon Lyons
Ian Maddieson
Carl Oberg
Stanisław Puppel
Ren Hong-Mo
Lloyd Rice
Mika Spencer
Diana Van Lancker
Bob Van der Veen
Karen Weiss
Andreas Wittenstein

*Anthony Davey, now at the University of Natal, Pietermaritzburg, was a Visiting Scholar at UCLA during the 1981-82 academic year, and contributed to Working Papers in Phonetics 54. His year at UCLA was supported in part by a Senior Research Grant from the Human Sciences Research Council of South Africa. This grant and the study leave granted by the Council of the University of South Africa is gratefully acknowledged by Dr. Davey and the UCLA Phonetics Lab. These acknowledgments were inadvertently omitted from WPP 54.

As on previous occasions, the material which is presented in this volume is simply a record for our own use, a report as required by the funding agencies which support the Phonetics Laboratory, and a preliminary account of research in progress for our colleagues in the field.

Funds for the UCLA Phonetics Laboratory are provided through:

NSF grant BNS-23110
USPHS grant 1 R01 NS18163-02
and the UCLA Department of Linguistics.

Correspondence concerning UCLA Working Papers in Phonetics should be addressed to:

Phonetics Laboratory
Department of Linguistics
UCLA
Los Angeles CA 90024
(U.S.A.)

The limits of biological explanations in phonetics

Peter Ladefoged
University of California, Los Angeles

[To be part of Proceedings of the Tenth International Congress of Phonetic Sciences]

In the preceding paper Lindblom has proposed broad goals and a wide scope for phonetic inquiry. He has suggested that the study of language should be firmly anchored within principles that apply in other fields. In this paper I am going to stress the uniqueness of language, and suggest that many (perhaps most) of the interesting things in the phonologies of languages cannot be explained in terms of principles that apply in other fields. I will provide illustrations of the kinds of things that show the uniqueness of our field. I will do this by making three points concerning phonetic differences among languages, all of which combine together to support the notion of phonology (as opposed to phonetics) that Anderson (1981) has proposed. Firstly the sounds of individual languages differ from one another in many ways that are due to chance and custom. Such differences cannot be explained in biological terms. Secondly, different languages use different articulatory gestures to produce sounds that are acoustically and auditorily very similar. Regular, and statistically reliable, differences in articulatory habits of this kind offer further testimony to the limits of biological explanations of phonetic phenomena. Thirdly the coarticulations that occur between segments are governed by language specific rules. Some languages choose to make certain sounds in a fairly constant way, irrespective of the surrounding sounds; but other languages allow these same sounds to be very influenced by the contexts in which they occur. These differences between languages are again matters of chance and custom, and are not explicable in terms of any general phonetic principles. All three of these points illustrate the real disagreement between Lindblom and myself concerning the relation between phonetics and phonology, which I will discuss in the last part of this paper.

That the sounds of languages are different from one another is no news to phoneticians. For centuries traditional descriptions of languages have offered comments such as those of Yan Zhitui (circa 560), a learned Chinese scholar who noted the difference between the Chinese dialects that are now Mandarin and Shanghai, saying: (1) "In the South, where the climate and the natural environment are mild, people speak with a clear and raised voice ... while in the North, where there are big mountains and valleys, people speak with a lowered and muddy voice." Within the European tradition, a millenium after these observations, but still more than three centuries ago, Wallis (1653) claimed that "The English as it were push forward the whole of their pronunciation into the front part of the mouth, speaking with a wide mouth cavity, so that their sounds are more distinct. The Germans on the other hand, retract their pronunciation to the back of the mouth and bottom of the throat, so that they have a stronger and more forceful pronunciation."

My first point is that there is usually no biological phonetic explanation for these differences between languages.There have been attempts to explain them in terms of the basis of articulation. Thus French sounds have been said to have the tongue more convex to the roof of the roof of the mouth, in comparison with

English in which the tongue is said to be slightly concave Honikman (1965). There have even been claims that some of the differences between languages reflect an anatomical or physiological differences among the speakers of those languages Brosnahan (1957). And, in fact, it seems true that on a few rare occasions linguistic phonetic differences may be ascribed to something like a difference in the base of articulation, and this difference may have anatomical correlates. Disner (forthcoming) has shown that there is a consistent statistically significant, difference between the vowels of Italian and those of Yoruba that may be due in part to the shapes of the lips of Italian as opposed to Yoruba speakers. Each of these languages has a seven vowel system that can be represented by symbols such as /i,e,ɛ,a,ɔ,o,u/. Disner's charts of the formants of a group of a 25 Italian speakers and a group of 10 Yoruba speakers are shown in Figure 1. Each vowel is represented by an ellipse (solid lines for Italian and dotted for Yoruba), with its center at the mean for that vowel, its axes along the two principal components of the distribution of the points, and the radii corresponding to two standard deviations. Roughly speaking, therefore, we may say that the ellipses enclose 95% of the points corresponding to each vowel. The mean formant frequencies are similar in both sets of vowels, indicating that there are no overall differences in headsize between the two groups of subjects. But, with the exception of /i/ and to a lesser extent /e/, the second formant is lower for the Italian vowels than for the corresponding Yoruba vowels. There is also other data that show that, in the case of /i/, the third formant is much lower for the Italian vowel than for the Yoruba vowel. These differences are precisely those that one would expect if Yoruba speakers, on the whole, used a larger mouth opening than that used by the Italians. Rounding affects the third formant rather than the second for vowels like [i]; and in other vowels the second (as well as the third) formant is lowered by decreasing the lip aperture. Accordingly, in the case of these two languages, there may be a biological phonetic explanation for at least some of the differences in the formant frequencies between the two groups of speakers. The possibility of overall differences in mouth opening is certainly compatible with the apparent facial differences between speakers of Yoruba and Italian.

However as Figure 1 also demonstrates, these are far from the only differences between Italian and Yoruba. The most obvious differences are in the F1 dimension. The vowels of Italian are more evenly distributed than those of Yoruba, in which /e/ and /o/ are much closer to /i/ and /u/ than to /ɛ/ and /ɔ/ respectively. Lindblom has provided good reasons for expecting that languages in general will have their vowels fairly evenly distributed throughout the available vowel space. The uneven distribution of the Yoruba vowels may be attributed to historical facts concerning the way in which the vowels of the original 9 or 10 vowel system have merged to produce the current 7 vowel Yoruba system. But, nevertheless, this historical explanation does not disguise the fact that present day Yoruba chooses to have vowels that are not evenly distributed. It chooses to resist the biological pressures that undoubtedly exist, and undoubtedly account for phonetic phenomena that are widespread in the languages of the world.

My second point concerns another kind of difference between languages, which demonstrates even more clearly that these differences are subject to the whims of fashion, rather than the rule of phonetic law. Some articulatory gestures can produce exactly the same sounds - the same acoustic structures - as other quite different movements of the vocal organs; and some languages habitually use one way of producing a given sound, and others another. The evidence on this point is somewhat indirect, but its seems that some languages use one kind of lip gesture to produce a high back rounded vowel of the [u] type

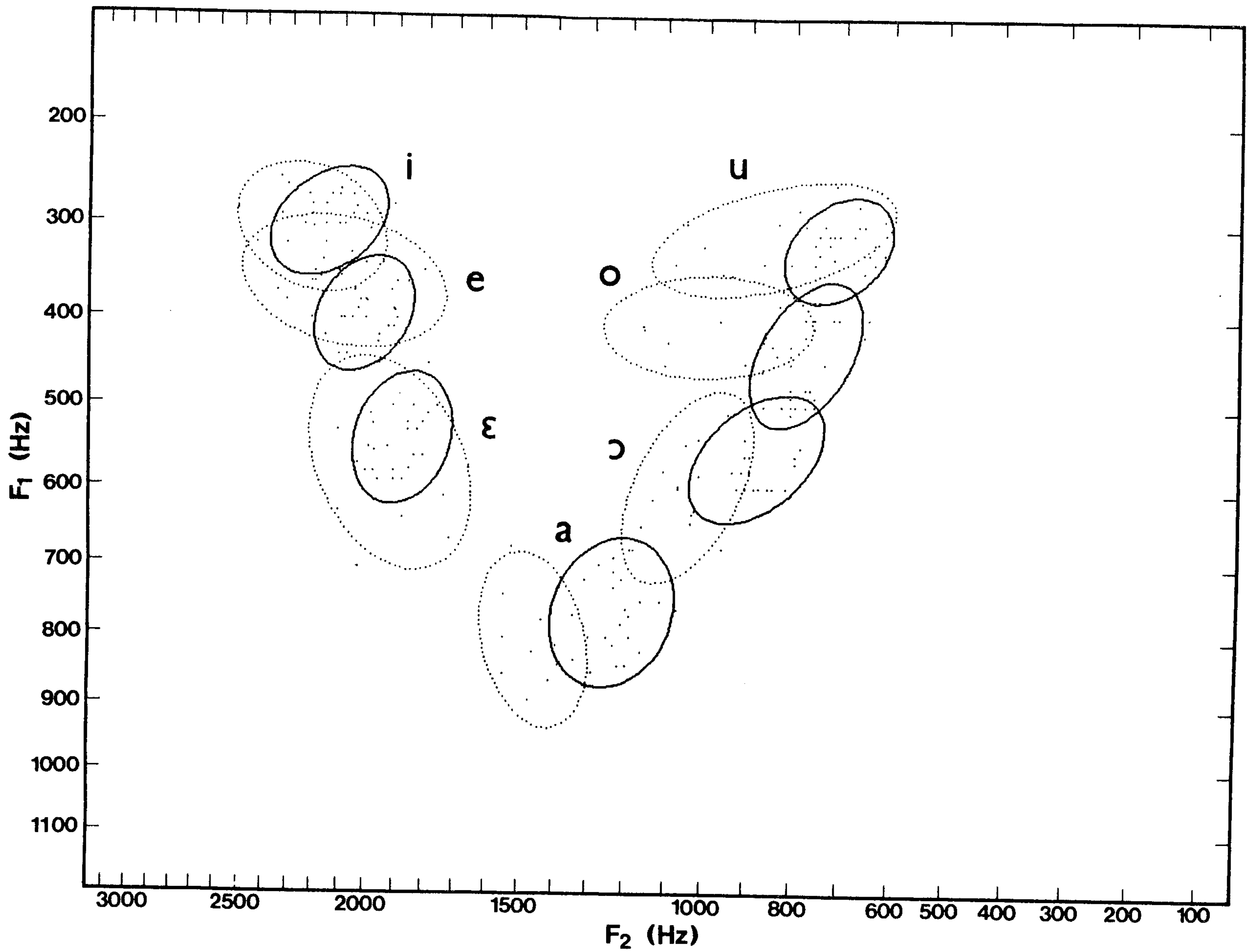


Figure 1. Formant plots of the vowels of 25 speakers of Italian (solid lines) and 10 speakers of Yoruba (dotted lines).

and others produce virtually the same vowel, using a different gesture, with less rounding of the lips being compensated by other vocal gestures such as more lowering of the larynx.

Linker (1982) photographed the lip positions of the vowels in the speech of eight speakers of Cantonese and eight speakers of French. She examined the correlation between formant frequencies and lip position in each of these (and other) languages, and showed that there are reliable, statistically significant, differences in the articulatory-acoustic relations. As she says, "if ... a (mean) speaker of Cantonese wanted to produce an [u] with a given set of formant frequencies, he would have considerably less horizontal opening than a speaker of French producing the same vowel." She concludes "These results indicate that languages differ greatly in the lip gestures they use to make the same acoustic distinctions among vowels."

Presumably children learning languages do so by watching as well as listening. Children of each group see the lip positions that are typical in their language and learn to make these sounds in that way. There may be a biological explanation for this difference between French and Cantonese, and it may be associated with the different anatomies of the different racial groups; but it seems unlikely. It is much more likely to be a case of language specific behavior that can be ascribed only to the whims of fashion.

There is certainly no anatomical basis for another case in which visual cues during language acquisition are probably the cause of adults having very different articulations when producing acoustically very similar sounds. As has been shown by Ladefoged (1979) a considerably higher proportion of American English speakers use an interdental fricative, as compared with the dental fricative which is more common in British English. In Californian English 75% of the speakers protrude the tongue between the teeth when saying words such as "thief", whereas virtually no speakers of British English make the fricative in this way. Again, there are no biological reasons for this. It is just that British English prefers to do things one way, and American English another.

My third point concerns coarticulation. We all know that in English, as in many other languages, the place of articulation of a stop is affected by the articulation of the following vowel. Thus the velar stop in "keep" has a more forward articulation than that in "cop." But to what extent is this difference in the stops <u>explained</u> by claiming that there is coarticulation with the following vowel? This would seem to imply that it is a necessary thing that languages have to do. If this were so, it would be nice, not only from the point of view of making biological explanations, but also from the point of view of simplifying the theory of phonology. Life would be easier if there were a set of universal, language independent, rules that took a string of segments specified in terms of a narrow, allophonic, transcription (or a matrix of distinctive features) and converted it into a sequence of continuously varying parameters of the kind required for a complete description (synthesis) of the utterances represented by the string. But for many years it has seemed to me to be unlikely to be the case. There is no force compelling languages to have a given degree of coarticulation; and, as indicated in Ladefoged (1972), it seems likely that French and English differ in the ways in which coarticulations occur. The French velar stops in <u>pique</u> [pik] and <u>Pâques</u> [pɑk] coarticulate more with the preceding vowel than the corresponding stops in the English words <u>peak</u> [pik] and <u>pock</u> [pɑk].

Recently Nartey (1982) has given a more rigorous demonstration of the language dependent nature of coarticulation. He recorded (among other data) eight speakers of Amharic and eight speakers of Navaho. Both languages have, among other fricatives, two sibilants which may be transcribed /s ʃ/. They also have similar vowels that may be transcribed /i,a/. The fricatives in each language were recorded in the context /i-i/ and /a-a/. The acoustic spectra at the midpoint of each fricative was determined using the UCLA WAVES computer analysis sytem (Wittenstein & Rice, hopefully forthcoming). These spectra were then convolved with an auditory filter as suggested by Bladon and Lindblom (1981) to produce representations of the auditory information.

The upper part of figure 2 shows the mean auditory spectrum of Amharic [ʃ] in the two vocalic contexts; the lower part shows Navaho [ʃ] in similar contexts. There are differences in the sharpness of the spectrum in the two languages. But, over and above this, it may be seen that the Amharic sounds show greater effects of coarticulation with the vowel, the spectrum having a noticeably higher peak in this Bark scale representation when in the context of [i] than when in the context of [a]. In Navaho there is very little coarticulation between this fricative and the adjacent vowel. Again we see that languages differ from one another in ways that cannot be predicted.

Once we have accepted the fact that languages differ fron one another in unpredictable ways we can go on to consider the really fundamental difference between Lindblom and myself. I find it necessary to distinguish three separate entities: the speaker, the listener, and the language. Various things follow from this distinction. The major point that I want to concentrate on is that I do not think that we can derive "the fundamental units and processes of linguistic structure deductively from independent premises anchored in physiological and physical realities" (Lindblom 1981). Instead I agree with Anderson (1981), and other generative phonologists who consider the interesting linguistic observations to be those that can be made in formal terms, independent of any other data. I disagree with scholars of the latter group only when they try to relate these linguistic observations to observable mental capacities. Instead I would rather describe properties of a language -- an observable social institution -- without having to consider what goes on in people's minds. A linguistic theory is a description of institutional norms, not of mental states. But this is a very small difference in comparison with the difference between generative phonologists as a group and scholars such as Lindblom who wish to view language as something rooted in physiological principles that are not specially related to linguistic structures. Whether language is an organ in the mind, as Chomsky (1975) puts it, or a social institution as I prefer to regard it, it has some properties of its own that are not dependent on our muscular abilities, or our auditory systems, or any cognitive mechanisms that apply to our general, non-linguistic, abilities. I do not, of course, wish to deny that these other factors have a great influence on language. But they should not be regarded as the only determinants of linguistic structure.

We should note here that the mental nature of language has been somewhat misleadingly presented by Chomsky. His notion that language is an organ of the mind is not very helpful. It is somewhat like saying that digestion is an organ of the body. Digestion is an ability that involves many components, including some things that are normally called organs such as the liver and the pancreas, as well as a number of other things such as saliva, mastication, and bowel movements. Digestion is like language in that it is a system rather than an organ in the usual sense. But a much better system for comparison with language is

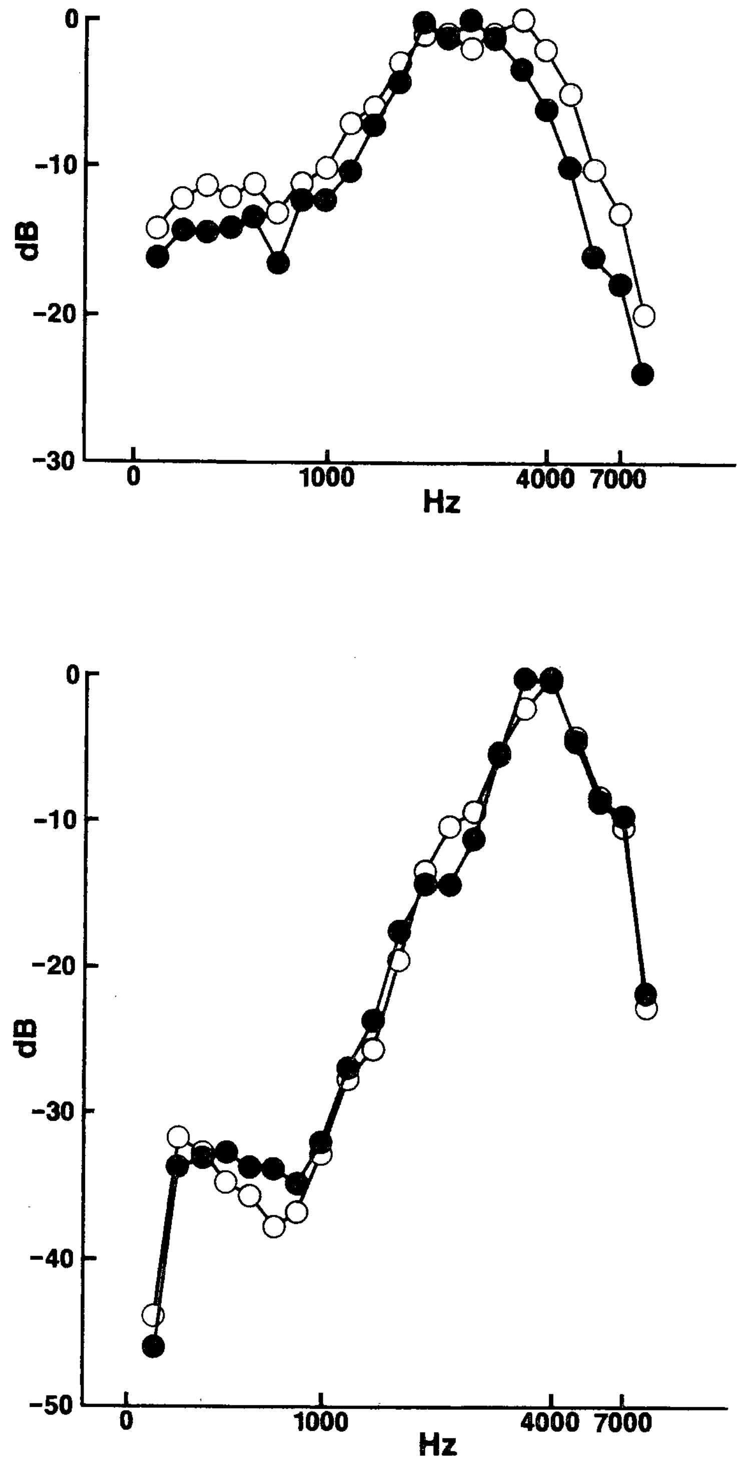


Figure 2. The mean auditory spectra of [ʃ] in the context [i-i] (open circles) and [a-a] (solid points). Data for 8 speakers of Amharic in the upper part of the figure, and for 8 speakers of Navaho in the lower part of the figure.

morality. Both language and morality can be, to a limited extent, localized within a particular side of the brain. Language is more on the left side, and morality on the right, as can be seen from the fact that patients with damage to the right side of the brain are more likely to perform anti-social acts than those with damage to the left side. Morality is a system of value judgments that form a moral code. This moral code is clearly a product of society; it is strongly influenced by the surrounding culture. But it is also a property of the individual, at least to the extent that the individual can choose to perform moral acts. Morality, or at least the capacity for performing moral acts, may even be like language, or the capacity for language, in being innate. Certainly one way to think of morality is as an over-developed herd instinct -- a self-organized, innate, drive for the preservation of the herd rather than the individual. An actual moral code is like a particular language in yet another way, in that it reflects the need for a balance between two conflicting forces: the desire for maximum individual freedom, and the necessity for social justice. These two forces might be likened to two conflicting forces in a language. The desire for individual freedom is like the wish of the speaker to be as lax as possible in conforming to the linguistic code -- an ease of articulation principle. The necessity for agreed social justice corresponds to society's need to communicate using an agreed set of distinct sounds -- a perceptual principle.

Putting this in terms of my argument with Lindblom, we should (as he says) always try to explain as much as possible about languages in general physiological terms, just as philosophers and anthropologists try to explain why particular moral codes prescribe some acts rather than others. But moral codes, like languages, rapidly become ritualized, and there is often no discernible reason why a particular society regards one thing rather than another as being moral. We should look for explanations, but we may not be able to find them.

Once the autonomous nature of at least some aspects of language is recognized, we can see even more clearly the importance of distinguishing descriptions of speakers and listeners from descriptions of languages. The units of abstract linguistics -- things such as phonemes and features -- are of little relevance for speakers and listeners. People organize their talking and listening mainly in terms of much larger chunks of sounds. Sometimes, as phoneticians, we are concerned with characterizing the behavior of listeners and speakers. When we do this we should probably not start with discussions of units such as phonemes. It is only the language that has phoneme size abstract units; when we are discussing the behavior of languages, units of this size are very important.

Lindblom's notions on pattern formation in languages are interesting in this respect. I agree that many of the patterns that we observe are indirect consequences of interacting systems, rather than primary components of the human behavior that occurs when they are produced. The analogy between the description of languages and the description of termite structures is very appropriate at this point. Phonemes may be like the arches in termite nests, visible to outside observers, but having no meaningful role in the activity of the individuals producing them. Speech appears to be composed of sequences of segments because of the interactions of the different systems of which it is composed. The complex gestures involved in producing syllables have diverse parts that look as if they are categorically distinct. We call these diverse parts vowels and consonants, but we must always remember that these are just names for readily distinguishable aspects of the stream of speech. Those of us who have been exposed to an alphabetic tradition may be influenced so that we are very conscious of the possibilty of describing speech in terms of units of this kind. But illiterates may have little or no concept of speech segments (Morais et al, 1979) Similarly those involved in adult literacy campaigns report that the concept of the segment is far from self evident. Intelligent adults who have been taught to write a few words cannot perform tasks such as naming other words that begin with the same segment (Jackson 1982).

We can carry the termite analogy a step further still. Just because the individual termite cannot be considered responsible for the design of the arches and pillars in a termite nest, it does not follow that it is not interesting to describe these pillars and arches. They are a necessary part of the termite nest; when a pillar or an arch is needed to support the edifice which the community requires, if the nest is to be one which survives, then the requisite structure will be present. In that sense, a language gets the phonemes that it needs. The segments and phonemes are present in the structure of that abstract entity, the language, considered as a social institution. Indeed, as Halle commented several years ago, "Almost every insight gained by modern linguistics from Grimm's law to Jakobson's distinctive features depends crucially on the assumption that speech [or, in my terms, language] is a sequence of discrete entities." (Halle 1964.) Halle, of course, has a different view of language to that being advocated here, in that he regards linguistic structure as having some kind of mental existence. But whether this is so or not, the point on which Halle and I agree, is that language is not, as Lindblom would have it, anchored in our _general_ phsysiological attributes. Language is something special.

It seems to me that the lessons of social evolution are clear, and are not as Lindblom would express them. Language has, to a great extent, become its own thing, and is largely not explicable in terms of outside forces. With the possible exception of codes of morality, it is the most developed and intricate of all social institutions, and now must be considered _sui generis_. In its evolution it has fed upon itself, and can be explained only in terms of unique principles. The current explanations may not be correct. Descriptions of languages in terms of the fashionable metrical phonology (Halle and Vergnaud, 1980) are, after all, at least superficially very different from those of the older generative phonology (Chomsky and Halle, 1968). But the thread of the uniqueness of language that runs through them is still valid.

There are therefore two reasons why we should not concentrate on examining linguistic phenomena in the way that Lindblom advocates. Firstly, language is, at least in part, its own thing and cannot be explained in terms of other principles. Secondly, one of the main tasks expected of phoneticians is to provide good descriptions of the sounds of languages. Historically this is how many of us have earned our bread and butter. It is often of great practical importance to be able to document how the sounds of one language differ from those of another. Lindblom stresses the similarities among languages, while I emphasize the differences. Both are worthy things to do. I agree that it is very important to investigate the fundamental bases of Language (with a capital L); but we must also note that languages (with a small l) are not all the same. There are vast differences among phonetic structures and many of these have arisen precisely because there are limits on the biological forces that affect the sounds of speech. This does not mean that these forces are not interesting. It is certainly part of the phonetician's job to reveal the general phonetic tendencies that account for many of the phonological regularities in the languages of the world. But we must also note that some of the really interesting linguistic phenomena are precisely those for which there is no general phonetic explanation. Languages often choose to be different because of the whims of fashion.

Phoneticans must be able to document these differences without expecting to explain them. We should always seek the larger verities, but we may not be able to find them. The individual termite will never know the meaning of the arch.

Footnote

1. I am endebted to Mr Chen Ping for showing me this quotation and for translating it for me.

References

Anderson,S. (1981). Why phonology isn't "natural". Linguistic Inquiry 12 493-589.

Bladon, A. and Lindblom, B. (1981). Modeling the judgment of vowel quality differences. J. Acoust. Soc. Amer. 69.5, 1414-1422.

Brosnahan, L. (1957). Genes and Phonemes. Ibadan: Ibadan University Press.

Chomsky, N. (1975). Reflections on Language. New York: Parthenon Press.

Chomsky, N. and Halle, M. (1968) The Sound Pattern of English. New York: Harper and Row.

Disner, S. (1983). Vowel quality: the contribution of language particular and language universal factors. UCLA Working Papers in Phonetics 58.

Halle, M. (1964). On the bases of phonology. In: The Structure of Language (Fodor and Katz, eds), Englewood Cliffs, New Jersey: Prentice Hall, 604-612.

Halle, M. and Vergnaud, J-R. (1980). Three dimensional phonology. J. Linguistic Research 1.1, 83-105.

Honikman, B. (1965). Articulatory settings. In In Honour of Daniel Jones (Abercrombie et al, eds), London: Longmans, 73-84.

Jackson, E. (1982). Personal communication.

Ladefoged, P. (1972). Phonetic prerequisites for a distinctive feature theory. In Papers in linguistics and phonetics to the memory of Pierre Delattre. (Valdman, ed), The Hague: Mouton, 273 -285.

Ladefoged, P. (1979). Review of J.C. Catford: Fundamental Problems in Phonetics. Language 55.4, 904-907.

Lindblom, B. (1982). Economy of speech gestures. In: The Production of speech (MacNeilage, ed): New York Springer Verlag.

Linker, W. (1982). Articulatory and acoustic correlates of labial activity in vowels: a cross-linguistic study. UCLA Working Papers in Phonetics. 56, 1-154.

Morais, J. et al. (1979). Does awareness of speech as a sequence of phones arise spontaneously? Cognition, 7, 323-331.

Nartey, J. (1982). On fricative phones and phonemes. UCLA Working Papers in Phonetics 55. 1-181.

Smith, C. and Tager-Flusberg, H. (1980). The relationship between language comprehension and the development of metalinguistic awareness. Paper presented at the Fifth Annual Boston University Conference on Language Development.

Wallis, J. (1653) _Grammatica Linguae Anglicanae_ London.

Wittenstein, A. and Rice, L. (forthcoming) _The UCLA WAVES analysis system_.

Yan Zhitui (c560) _Family admonitions_ (annotated by Zhao Ximing). Shanghai Press Taiwan.

Non-quantal aspects of consonant production: a study of retroflex consonants

Peter Ladefoged
Phonetics Laboratory, Linguistics Department, UCLA

Peri Bhaskararao
Department of Linguistics, Deccan College, Pune, India

Some years ago Stevens and Blumstein (1975) stated: "The human vocal apparatus is capable of producing a wide variety of speech sounds. With the full range of articulatory manipulations" of which the vocal tract is capable, "it would seem that there should exist an infinite number of articulatory possibilities to be available for use in the language systems of the world. However, in fact, there seems to be a finite number of speech sounds in language characterized by remarkable uniformities in both their structure and function."

It is precisely this observation that we wish to challenge. Moreover, we will do so by reference to the same class of sounds, retroflex stop consonants, as that discussed by Stevens and Blumstein. It is not true that there is a "finite inventory of speech events in language" in the way that they suggest. We do not wish to maintain the exact opposite of their statement. There is not an infinite set of possible speech sounds; there are obvious physiological limitations and limits on the noticeable perceptual differences among sounds. But, as we will show, the inventory of speech sounds actually used in different languages is very much larger than that suggested by Stevens and Blumstein. When considering variation in place of articulation, we cannot find any evidence for the suggestion that there are "regions of acoustic stability _which may define_ the articulatory inventories used in natural languages" (Stevens and Blumstein 1975, our italics). We of course recognize that some places of articulation are far more favored than others. As Maddieson (1980) has shown, velar stops occur in 99% of the languages of the world, whereas uvular stops occur in only 14%. Similarly virtually every language has a dental or an alveolar stop, but palatal stops occur in only 18% of all languages. We believe that these differences are due to anatomical considerations rather than to quantal theory notions. In some sense which we cannot yet quantify it is easier to make a velar or dental or alveolar stop, than a uvular or palatal one. Similarly the comparative infrequency of retroflex stops (11% of the languages of the world) is due to the comparative complexity of the articulation required. Variations in the relative frequency of occurrence of different places of articulation are not due to a tendency to use "quantal" sounds.

Stevens and Blumstein go on to say "similar limitations on the possible set of universal phonetic categories also seem to be based on constraints of the perceptual auditory system". But at this point we should note that there is no necessary connection between the quantal theory and any particular constraints of the auditory system, if those constraints are presumed to be in the form of auditory property detectors. It would be quite possible for speakers to prefer so-called quantal sounds without the auditory system having particularized property detectors. Equally it would be possible for the ear to have developed detectors for certain prominent properties that distinguish speech sounds without there being a tendency for languages to have sounds that are acoustically stable despite a certain amount of articulatory variation. Neither theory adds support to the other. Each separately specifies a desirable feature that a speech communications system might be expected to have. It is just a pity that neither theory provides a viable way of accounting for variations in place of articulation.

We should now consider what would count as evidence in favor of a quantal theory of place of articulation. Obviously it is not sufficient to show that languages use only a small number of phonological oppositions. This is a phonological fact that has very little to do with articulatory phonetics. The reason languages use only a small number of contrasts is not because there is only a very limited set of articulations maximally suitable for use in a linguistic communication system. It is just that languages do not need a large set of contrasts. Human perceptual and cognitive systems require only a small number of distinct signals to encode all the words and sentences of a language. But these signals are not necessarily the same in every language. There are phonetic differences among the sounds that languages use for phonological contrasts.

More direct evidence is required to substantiate the notion of quantal places of articulation. One possibility for providing supporting evidence would be to use models of the speech production process to show that in some areas comparatively large changes of articulatory parameters cause only small changes in acoustic output. This is the technique used by Stevens (1972) in his early exposition of the theory with regard to vowels. Some places of articulation have also been discussed by Stevens (1972) who suggested that constrictions in certain regions of the vocal tract correspond to sounds that are acoustically unstable. These regions are said to form boundaries between the four more stable possibilities:- (1) "/g/ or /k/ before a back vowel, or the velar fricative /x/"; (2) "/g/ or /k/ before a front vowel, or the fronted velar fricative /c/"; (3) "the fricative consonant /š/"; and (4) "/s/". (Stevens 1972:63) This is a somewhat counterintuitive division of the continuum of possible places of articulation as far as linguists are concerned, particularly in its claim that there is a quantal difference between back and front velar consonants. Moreover it does not have any relevance to retroflex consonants. In their later work, "Quantal aspects of consonant production and perception: a study of retroflex consonants", Stevens and Blumstein (1975) do not report any additional modeling evidence in favor of their theory. This paper provides no data that in any way leads one to think that certain articulatory areas can be correlated with sounds with greater acoustic stability. With respect to variations in place of articulation there is very little reason to believe that "stable acoustic patterns can be generated by articulatory movements which do not have to be precisely executed" (Stevens and Blumstein 1975).

The quantal theory predicts that variation of an articulation might be permitted (or even expected) in the speech of a given speaker or the speakers of a given language. What it does not predict is that there might be comparatively stable, systematic, differences in articulation of a 'quantal' sound between two languages. It seems to us that it would be damaging to the theory if it could be shown that speakers of different languages do not have small variations in articulation, but instead maintain systematic differences within what Stevens and Blumstein would regard as a single quantal sound. Thus it would count as evidence against the quantal theory if it could be shown that languages do not have just dental, alveolar and retroflex sounds, at certain comparatively fixed places in the vocal tract, but instead exhibit a continuous range of places of articulation, some languages choosing certain areas within this range, and others choosing others. We propose to demonstrate that this is in fact the case by reference to retroflex sounds in different Indian languages. We will be concerned primarily with Hindi, an Indo-Aryan language spoken as a first language mainly in the Northern part of India, and Telugu, a Dravidian language spoken in Central and Southern India. We will also report some additional data on Tamil, another Dravidian language spoken in Southern India.

Several authors have noted differences among sounds classed as retroflex. For example, Firth (1957) notes differences between Brahmin speakers of the Satara dialect of Marathi and speakers of Northern Indian languages such as Hindi and Urdu. Svarný and Zvelebil (1955) indicate that there might be differences in the degree of retroflexion in Hindi, Telugu, and Tamil, on the basis of their x-ray studies of a single speaker of each language. Ladefoged (1964), using palatograms and linguagrams, points out that the degree of retroflexion in West African languages such as Ewe is not as great as that in many Indian languages. Bhat (1973), in a survey of retroflex consonants in 150 languages, notes that the point of articulation and the degree of curling of the tip of the tongue vary from language to language. Lastly, Catford (1977) observes that most retroflex stops (presumably including those spoken in the central and Southern parts of India) "are typically sublamino-prepalatal" (i.e. made with the underside of blade of the tongue touching the roof of the mouth well behind the alveolar ridge). He also notes: "However, in Hindi and some other North Indian languages the so-called ′retroflex′ stops are sometimes no further back than apico-postalveolar′ (i.e.made with the tip of the tongue just behind the alveolar ridge).

Previous studies of retroflex stops

Despite the wealth of opinions cited above, there have been very few experimental studies of retroflex consonants. The best and most detailed work has been done on Tamil. Svarný and Zvelebil (1955) made a lengthy series of palatograms, linguagrams and x-rays of a Tamil speaker, supplemented by similar experimental data on a speaker of Telugu and a speaker of what they call Hindustani of a kind that would now be called Urdu. They note that, as they have only a single speaker of each language, they cannot tell if the differences they found were speaker specific, or were real differences between languages. They found that the degree of reroflexion was greatest in their Tamil speaker, who regularly made [ṭ] with the tip of the tongue making contact in the prepalatal region. Both their Urdu speaker and their Telugu speaker had less retroflexion. Their x-rays also show that the degree of retroflexion is affected by vowel context, being less before a high back vowel. This fact makes some of their x-rays not entirely comparable to those we will be discussing below.

Palatograms and x-ray photographs are also given in the very detailed study of a single speaker of Tamil by Balasubramanian (1972). In this case the contact for [ṭ] definitely involved the underside of the blade of the tongue in the palatal region. Other articulatory studies include that of Dave (1977) who used palatograms to investigate his own pronunciation of Gujerati. He concluded that he used the underside of the tip of the tongue curled back into the palatal region when pronouncing [ṭ] and [ḍ]. Nihalani (1974) studied his own pronunciation of Sindhi. He notes that, in his speech, the sound is not "a true representative of the traditional category of sounds known as ′retroflex′ stops. The phonetic label ′apical post-alveolar′ would more adequately describe its formation." In addition Stevens and Blumstein cite a publication by Wierzchowska (1965), which we have not seen. The diagram of a retroflex articulation which they reproduce from this publication is not very similar to those which will be discussed below.

The acoustic results of retroflexion have been studied by a number of authors. The general concensus seems to be that retroflexion affects mainly the higher formants. Fant (1968) notes that a retroflex modification of alveolar sounds lowers F4 so that it comes close to F3; but a retroflex modification of palatal sounds modifies F3; so that comes close to F2. As a result of a theoretical analysis of the relative sizes of the cavities involved, Stevens and Blumstein (1975) note that "(1) the F2 transition is similar for both retroflex and non-retroflex consonants; (2) the F3 transition tends to be rising from the retroflex position; and (3) F4 starts very low, and undergoes a brief interval in this position before rising rapidly to the normal F4 position for the vowel." They go on to note that:"the overall acoustic pattern is characterized by a clustering of F2, F3 and F4 in a relatively narrow frequency region." The latter point is confirmed by Dave (1977). Stevens and Blumstein also find that retroflex stops have a burst of noise with a center frequency near F3 or F4. Dave (1977) notes that he has difficulty observing this, but, in general, he does find that the bursts for Gujerati dental stops are higher than those for retroflex stops. Ramasubramaniam and Thosar (1971) give a series of rules for synthesizing retroflex sounds, including: "when /ṭ,ḍ,ṇ/ [our symbols] precede and/or follow a vowel, the steady-state value of F3 of the vowel is lowered by about 225 cps." They also list the locus for both F2 and F3 in association with front vowels as 1900 Hz, with central vowels as 1450 Hz, and with back vowels as 1000Hz. Dave (1977) is somewhat critical of this work. On the basis of his acoustic analysis of two speakers he considers the F2 locus to be at 2100 Hz, with F3 and F4 having very low, but not fixed, loci. He also notes that in both his data and that of Stevens and Blumstein (1975) there are much greater formant transitions in going from a vowel into a retroflex consonant than in going from a retroflex consonant into a following vowel. Walldén (1974-75) observes that in Tamil "compared with a non-retroflex a postvocalic retroflex, especially ɻ and ɭ, has a clear lowering effect on the preceding vowel, and F3 is mainly responsible for this."

Procedure

Our general procedure was to compare the articulatory qualities of the retroflex stops as spoken by a group of Hindi speakers and a group of Telugu speakers. We considered groups of speakers of each language so that we could abstract away from the personal idiosyncrasies of the individual speakers, and determine the general characteristics that must be associated with each language. In addition we have data from a single Tamil speaker, which we are able to supplement with comparable data from the literature on two other speakers, so that we can report on a small, and possibly not representative, sample of that language.

Saggital x-rays were made of five Hindi speakers producing the voiceless dental and retroflex stops as in the words /ata/ 'is coming' and /ata/ 'flour'. In each case the speaker said the word twice and then, instead of saying the whole word a third time, stopped in the middle, holding the consonant position while an x-ray picture was taken. Speakers were carefully practiced in this procedure, so that we are fairly sure that the x-ray pictures provide a valid characterization of the articulations, except for the fact that many speakers immediately lowered the velum so that they could breathe through the nose. As a

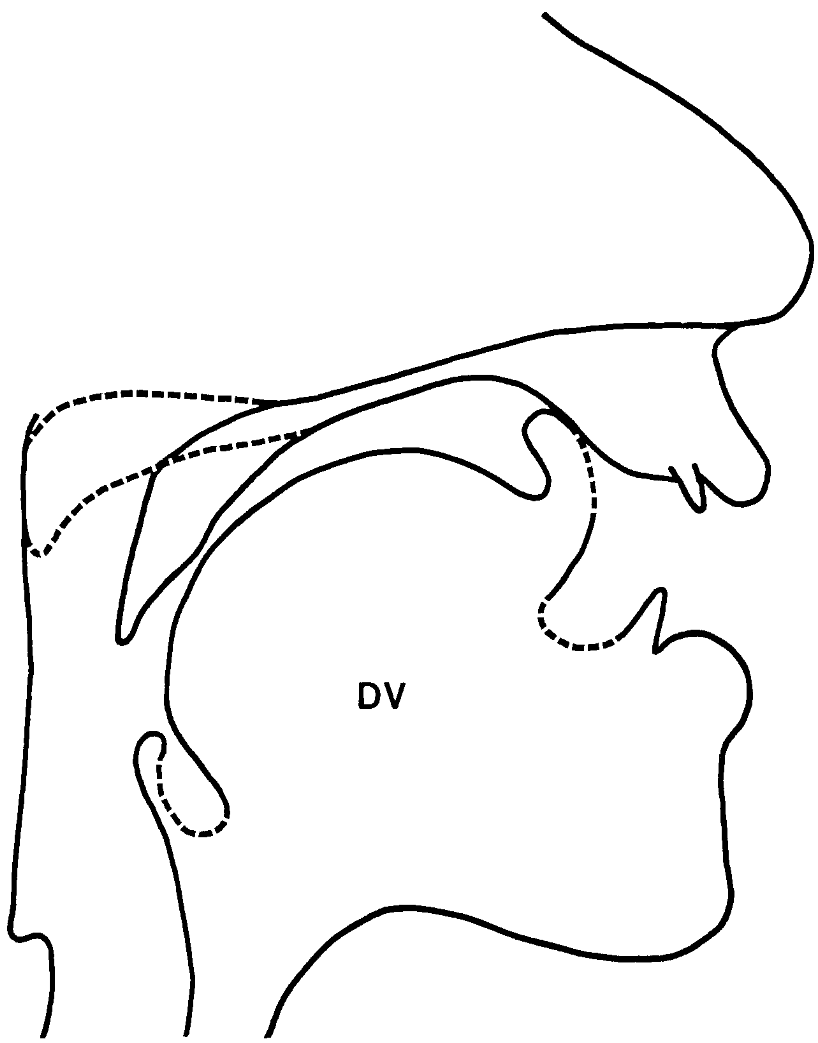


Figure 1. Example of a tracing from an x-ray of a retroflex consonant produced by a Telugu speaker. Solid lines indicate reliable data, dashed lines indicate data of varying reliability.

further check on the validity of the procedure, tape recordings were made of the utterances prior to the x-rays being taken. There was considerable background noise on the tape recordings, making them unsuitable for detailed acoustic analysis. But they were quite satisfactory for checking the naturalness of the pronunciations used. All five Hindi speakers appeared to be talking in a completely normal way, and the x-rays appear to have been taken almost immediately after the first vowel of the third utterance, at a time when the tongue might be expected to be in the appropriate position. In order to delineate further the exact articulations used, a stripe of barium was painted along the middle of the tongue. In addition, impressions of the oral cavity were made. These impressions were cut in half along the midline, so as to determine more precisely the shape of the anterior part of the mid sagittal section of the vocal tract.

Similar procedures were followed for a group of Telugu speakers. Five Telugu speakers, including two of those who were tape recorded, were x-rayed, using the procedures described above, except that the subjects said the Telugu words /paata/ 'old' and /paaṭa/ 'song' The x-ray data for one of the Telugu speakers had to be discarded. Judging from the notes taken during the x-ray procedure, and also from subsequent listening to the tape recording, it seems that he was unable to follow the instructions.

We presume that any factors such as the use of a formal reading pronunciation, or the use of a variant tongue position during the x-ray procedure, apply equally to both the Hindi and the Telugu speakers. Furthermore the consonants were embedded in similar vocalic contexts in both languages. Accordingly, any differences between the articulations of the sounds as pronounced by the two groups may be considered to reflect real differences between the two languages.

In addition to the groups of Hindi and Telugu speakers, we also obtained x-rays of a single speaker of Tamil, producing the Tamil words [vadai] 'torture' and [vaḍai] 'doughnut'.

Interpretation of the x-rays

As a precaution against bias in interpretation of the x-rays, a double blind procedure was used. The identifying marks on each x-ray film were covered up, and a new set of marks were added before the tracings of the mid-sagittal positions of the vocal organs were made. In this way the experimenter did not know the intended articulation or the language of the speaker when making the tracing.

The barium paste on the midline of the blade and front of the tongue was always clearly evident, so we are confident of the validity of our tracings of the most important (from the point of view of this discussion) of the articulators. The location of the corresponding points on the upper surface of the vocal tract, the midline of the hard palate and the alveolar ridge, was often more difficult to determine on many of the x-rays. But because we had dental impressions of each of the subjects we are also completely sure of our data in these areas. We were often less certain of the position of the underside of the blade of the tongue, particularly when it was obscured by the teeth. But we were usually able to infer the location of most of the underside of the tongue from the parts that were clearly delineated.

Our ability to make accurate tracings may be illustrated by figure 1, which shows the retroflex articulation of one of our Telugu subjects. The solid lines show reliable data, and the dashed lines indicate positions that vary from being reasonably reliable to being merely estimated on the basis of very little evidence. The extreme curvature of the upper surface of the tongue is clearly marked. We can also see, very clearly, the part of the underside of the tongue shown by the solid line, which appears between the teeth. As the tip of the tongue is well defined, the location of the part of the underside of the tongue that must be behind the upper teeth can be determined with reasonable confidence. But, as this part of the tongue is not actually visible, it is shown by a dashed line. The remainder of the underside of the tongue and the floor of the front of the oral cavity are totally obscured by the lower teeth. The dashed line in this area can be regarded as no more than a reasonable estimate. Note also that the velum was lowered at the time of the x-ray, as shown by the solid line. But as this sound is plainly a stop, we can assume that its position must have been much as is shown by the dashed line. It should be noted that none of our conclusions are based on the dashed lines shown in this or any other figure.

Results

Tracings of the retroflex consonants as pronounced by the five Hindi speakers are shown on the lefthand side of figure 2. The retroflex consonants for the four Telugu speakers are shown on the righthand side of the figure. It is difficult to devise an appropriate quantitative measure of retroflexion. Our nearest approach is to estimate the angle between the mean slope of the surface of the blade of the tongue and that of the front of the tongue. Estimates of these angles are shown in the lower part of each tracing. It is quite clear that four out of the five Hindi speakers have less retroflexion than any of the Telugu speakers. Only the fifth Hindi speaker (BNM, at the bottom left of the figure), has a considerable degree of retroflexion.

As noted, we also have x-rays of a single Tamil speaker; and there are comparable x-rays of the retroflex stops of two other Tamil speakers published in the literature. Svarný and Zvelebil (1955) reproduce x-rays and tracings for a Tamil speaker producing the stop in [aṭa] (hey!). Balasubramanian (1972) reproduces a photograph of an x-ray of his pronunciation of the retroflex stop in [pa:ṭI] (grandmother). The vocalic context in this case is thus a little different from that in the other data being reported here, in which the stop was always before the vowel [a]. But the palatographic data which Balasubramanian reproduces for other words containing retroflex stops show that the articulation for the stop in the x-ray is very comparable with (almost indistinguishable from) that in words before [a]. In addition, Svarný and Zvelebil's paper includes data on a speaker of Urdu, a language very similar to Hindi, heavily influenced by Persian, and a speaker of Telugu. The Urdu speaker was producing the retroflex stop in [ḍab] (purse), which makes it comparable with our data. But unfortunately the x-ray photograph and tracing for the stop pronounced by the Telugu speaker are not comparable with our data, as the stop occurred before a high back vowel in [paḍu] (to appear).

The articulations for the three Tamil speakers and the Urdu speaker are reproduced in figure 3. It may be seen that all three Tamil speakers are more like our Telugu speakers than our first four Hindi speakers. This is not surprising as Tamil is a Dravidian language very closely related to Telugu. The Urdu speaker is fairly similar to the majority of the Hindi speakers. Again this

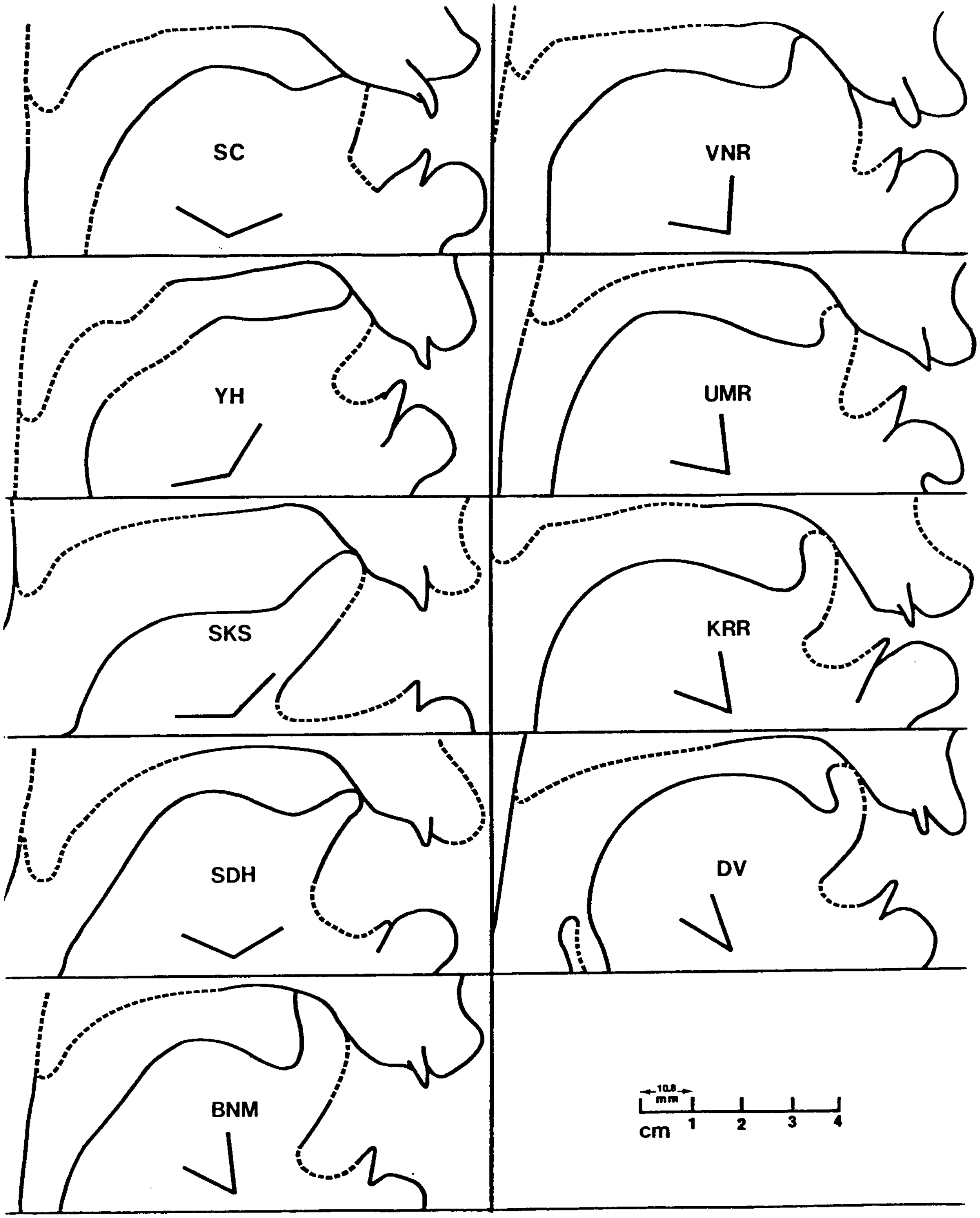


Figure 2. Retroflex consonants in Hindi (left hand side) and Telugu (right hand side).

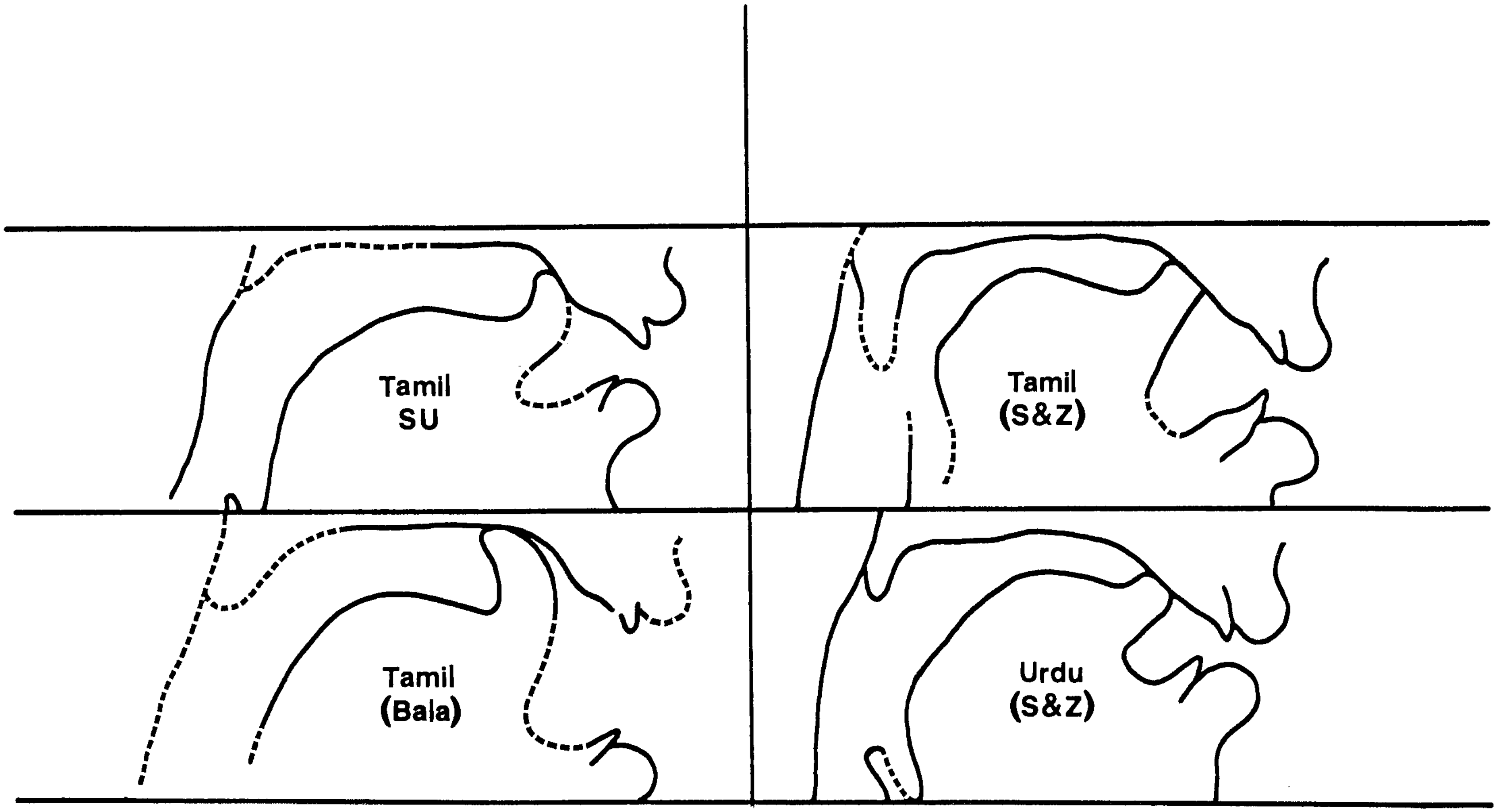


Figure 3. Retroflex consonants in Tamil and Urdu. Subject SU was recorded as discussed in the text. The subjects labeled (S&Z) are from Svarný and Zvelebil (1955). The subject labeled (Bala) is in Balasubramanian (1972).

is not surprising considering the closeness of Hindi and Urdu (which may even be regarded as dialects of the same language).

Table 1 summarizes the x-ray data in terms of measurements of the angle of retroflexion. With such small numbers of subjects, and data that do not form a normal distribution, it would be unwise to attempt to make any statistical generalizations. We will simply note that all seven speakers of the Dravidian languages Tamil and Telugu have retroflex stops with a high degree of retroflexion, and that all except one of the six speakers of the Northern Indian language Hindi/Urdu have a smaller degree of retroflexion.

Table 1. Measurements of the angle of retroflexion for the six speakers of the Northern Indian language Hindi/Urdu, and the seven speakers of the Dravidian (Southern Indian) languages Telugu and Tamil. The angles are measured in degrees.

Language	Speaker	Angle	Language	Speaker	Angle
Hindi	SC	132	Telugu	VNR	84
Hindi	YH	130	Telugu	UMR	71
Hindi	SKS	135	Telugu	KRR	61
Hindi	SDH	126	Telugu	DV	38
Hindi	BNM	56	Tamil	SU	95
Urdu	(from S&Z)	126	Tamil	(from S&Z)	103
			Tamil	(from Bala)	61

Discussion

It seems that, in general, Hindi/Urdu speakers have a different target for their retroflex stops than the speakers of the two Dravidian languages. The question then arises as to whether these different targets can be represented by points on the same continuum, or whether they are simply both 'quantal' sounds. It is possible that instead of calling all these sounds retroflex we should have recognized that there are two distinct sound types, which we might, following Svarný and Zvelebil (1955), call retroflex (when the underside of the tip of the tongue is involved) and cacuminal (when the tip makes contact with the post-alveolar ridge as in most Hindi sounds of this type). But this does not resolve the quantal issue. It is not that there are two, and only two, possible gestures for retroflex consonants. Instead there is a continuum going from a very retroflex sound of the kind used by speakers of Telugu and Tamil, through a slightly retroflex sound of the kind that occurs in Hindi, to a non-retroflex sound involving the tip of the tongue such as English alveolar [d]. The Telugu sounds made with the underside of the tip of the tongue may be at or near the endpoint of this continuum. But the Hindi speakers use an articulation somewhere near the midpoint of the continuum. They choose to make their stops just slightly behind the alveolar ridge, so that they are not like the Telugu stops, but nor are they like the lingual stops in other languages such as English. There is no way we can get around the fact that these Hindi stops have a slight degree of retroflexion -- a very non-quantal notion.

It is interesting at this point to consider how Chomsky & Halle (1968) consider this problem. They comment on Ladefoged (1964) as follows:

"Ladefoged notes that the Ewe retroflex consonant "sounds slightly different from the retroflex stop found in Indian languages such as Hindi" (p.18). If this difference is systematic, it would clearly have to be reflected in the grammars of these languages. It is, however, quite sufficient to note that the point of contact between the tongue and the roof of the mouth is somewhat more advanced in one language than the other. This fact would presumably be reflected in low-level phonetic rules that assign numerical values to the different features. The existence of a systematic phonetic difference does not, therefore, in itself constitute a necessary and sufficient condition for postulating an additional point of articulation."

But it does, of course, tend to invalidate the quantal theory.

Proponents of the quantal theory could still point out that we might be overlooking some property that caused us to consider both the Hindi and the Telugu sounds to be retroflex in the first place. There might, after all, be some acoustic property common to both these sounds. This is an empirical issue. All we can say at the moment is that we cannot find an acoustic property common to all retroflex sounds. In fact, we think it very unlikely that there is any property that could be recognised by a "retroflex detector" that is part of our human sensory apparatus. In this paper we will not be reporting any new experiments on the perception of dental and retroflex consonants. Accordingly, we will consider only briefly the notion of auditory property detectors. Stevens and Blumstein suggest that listeners identify sounds such as [t,t,k] by means of two auditory property detectors. One of these detects a primarily falling onset of the spectrum, which they associate with the burst of [t] in relation to the following formants. The other detects a spreading spectral pattern which they associate with the burst and formant transition for [k]. They speculate that [t] is a more complicated (and therefore linguistically more marked) sound because it has to be identified by using a combination of these two property detectors.

It is difficult to find any evidence in favor of the existence of auditory property detectors of the kind proposed. Stevens and Blumstein tested the perception of 8 speakers of Indian languages listening to synthetic speech stimuli varying along an acoustic continuum going from [t] through [t] to [k]. They make no claim that these sounds could be produced by human speakers. From an articulatory point of view it is possible to make a continuously changing set of sounds from an apical dental [t̪] to a retroflex [t], moving the articulatory contact made by the tip of the tongue farther and farther back in the oral cavity. It is also possible to make another smoothly changing set of sounds, starting from a laminal dental [t̪] and moving the contact made by the surface tongue farther and farther back until a [k] (or an even more retracted) sound is produced. Neither of these continua is much like the sounds used by Stevens and Blumstein. We are not impressed by the fact that the listeners' responses to their synthesized, articulatorily impossible continuum, followed the usual kind of categorical perception behavior that has been reported in many previous studies. In our view categorical perception of these synthetic stimuli is not sufficient evidence for feature detectors. Nor,as we have noted, does it have any bearing on the quantal theory.

If there is no acoustic property that distinguishes all retroflex sounds from other sounds, we must consider yet again why they are all classified together. The most probable reason is that they have a family resemblance in the Wittgensteinian sense (Wittgenstein 1952). In the early stages of scientific inquiry objects are often grouped together because of a family resemblance. In such a grouping A may be like B, B like C, and C like D. But A and D may have nothing in common. If this is the case, then we may be superimposing the wrong notion of what constitutes a family on the data. A, B, C, and D may not be truly related at all.

If we carry this reasoning a step further, it follows that there is not a specific category "retroflex" but merely an ill-defined set of sounds that merges into neighboring sets of sounds, such as "alveolar" which are also not well defined. This means giving up the notion that the various places of articulation can be divided into distinct categories just as many phoneticians have given up the notion of dividing the vowel spaces into distinct categories labeled "high," "mid," and "low"). One might hope that the notion of distinct phonetic categories of place of articulation could be salvaged in some way by reference to phonological constraints. Thus one might make it a definong characteristic of a group of sounds being considered a single place category that, phonologically, languages did not use differences among these sounds to form contrasts. But we do not know any way in which we can know what sound differences <u>could</u> be contrastive in a language. The Hindi sounds are perceptually different from the Telugu sounds (as is demonstrated not only by the descriptions in the literature which we have cited, but also by the fact that Telugu children learn to make the one tupe of sound and Hindi children the other). This perceptual difference could conceivably be used contrastively within a single language. The situation with places of articulation seems to be similar to that with vowel heights. We can no more give a phonetic definition of what is meant by retroflex than we can give a definition (i.e. one that doesn not mention the sounds of a particular language) of what is meant by a mid-vowel.

Although we are unable to formulate phonetic criteria for deciding whether differences between sounds could be used phonologically within a language, we must note that the articulatory differences between languages are smaller than those commonly used to distinguish words within a language. Thus although there are differences between Hindi [ḍ] and Telugu [ḍ], and between Hindi [ḍ] and English [d], we do not know of any language that uses either of these pairs of sounds to form phonemic contrasts. (although of course, Malayalam does have a three way contrast between a dental [d̪], an alveolar [d] similar to the English sound, and a retroflex [ḍ] similar to the Telugu stop). In general, we suspect that the need to maintain a sufficient auditory distinction between contrasting sounds may result in languages usually having comparatively large acoustic differences between phonemes. But differences between languages (and between accents of the same language) can be maintained by references to small details of sounds that are manifested over longer stretches of speech. The need to communicate puts fairly rugged demands on a system. The maintenance of one's sociolinguistic identity can be achieved by subtler mechanisms.

Considering the complete set of data presented in this paper, it is evident that there is a continuous range of possible articulations. Because languages need only a limited number of sounds, and because sounds need to be sufficiently distinctive, only a few points in this range are used within any one language. We should emphasize once more that we are well aware that the languages of the world favor some points in this range rather than others. But this is not because these sounds are in any sense quantal; it is usually just that they are easier to make. We should also make clear that we fully appreciate Stevens and Blumstein's point that some comparatively large changes in articulation produce comparatively minor acoustic changes. Indeed, it seems likely that the Telugu articulations involving the under side of the tongue in various degrees all have very similar acoustic consequences. But the Hindi retroflex stops do not seem to be in a special region.

We can now see how it is that there is not a small set of sounds that the languages of the world use to form contrasting units. Each language may use only a small number of sounds. But they may choose these sounds from a very wide range of possibilities. We do not think it profitable to call this an infinite set of possibilites. But there are several different continua and languages are free to choose points anywhere within each of these continua. The findings reported above will come as no surprise to anyone who has looked at phonetic data in a wide variety of languages. There are a number of cases in which it seems evident that the sounds of one language are not identical with sounds that may be given similar classificatory labels in other languages. The data reported above show that,rather than there being a simple category retroflex, there are degrees of retroflexion just as there are degrees of vowel height. As shown by Disner (1979), languages may have vowels at virtually any point within the vowel continuum. There are no preferred vowel qualities with the possible exception of [i,a,u] which are preferred not because they are "quantal" vowels in which comparatively large variations in articulation do not have much acoustic effect, but because they are end points within the articulatory area, and maximally distinct. We are confident that we could have produced data similar to the retroflex data by considering, for example, dental nasals in Marathi, in which the tongue tip touches the back of the upper incisors, in comparison with the similar sounds in Malayalam, in which the tip of the tongue is protruded between the teeth. Or, we believe, we could have shown measurable differences between the murmured nasals in Sindhi (which have a comparatively high airflow) and those in the Southern Bantu language Tsonga (which are more voiced than breathy). Other examples of quantifiable variations between similar sounds in different languages have been given elsewhere (Ladefoged, 1980, Disner 1970, Lindau-Webb 1980, 1982, Nartey 1982).

Given the reliable, significant, measurements, and the documented observations of phoneticians who have made similar observations for many years , it is difficult to see why anyone should still believe in a quantal theory that predicts that there should be "a finite inventory of speech events in language" in the way suggested by Stevens and Blumstein (1975). Perhaps the primary reason is that it would have been nice if this theory had been true.

Acknowledgements

Our thanks are due to Professor Bh. Krishnamurthi for inviting us both to visit his department at Osmania University, and providing us with facilities as well as much helpful advice. We are also grateful to Dr Paroo Nihalani for arranging for us to use the Phonetics Laboratory at CIEFL, Hyderabad, and, most particularly, obtaining the x-ray data for us. Wise comments and helpful references were provided by John and Manjari Ohala, Aditi Lahiri, Pat Keating, and Ian Maddieson.

References

Balasubramanian, T. (1972). The phonetics of colloquial Tamil. Ph.D. thesis, University of Edinburgh.

Bhat, D.N.S. (1973). Retroflexion: an areal feature. Working Papers on Language Universals 13, 27-67.

Bhat, D.N.S. (1974). Retroflexion and retraction. Journal of Phonetics 2, 233-8.

Catford, J.C. (1977). Fundamental problems in phonetics. Bloomington and London: Indiana University Press.

Chomsky, N. and Halle, M. (1968). The Sound Pattern of English. New York: Harper and Row.

Dave, R. (1977). Retroflex and dental consonants in Gujerati: a palatographic and acoustic study. Annual Report of the Institute of Phonetics, University of Copenhagen 11, 27-156.

Disner, S. (1980). Evaluation of vowel normalization procedures. J. Acoust. Soc. Amer 67.1, 253-261.

Fant, C.G.M. (1968). Analysis and Synthesis of speech processes Manual of Phonetics (ed. by B. Malmberg) Amsterdam: North Holland.

Firth, J.R. (1957) Word palatograms and articulation. in Papers in Linguistics 1934-1951. London: Oxford University Press.

Ladefoged, P. (1964). A Phonetic Study of West African Languages Cambridge: Cambridge University Press.

Ladefoged, P. (1980). What are linguistic sounds made of? Language 56.3, 485-502.

Lieberman, P. (1977). Speech physiology and acoustic phonetics. New York: Macmillan.

Lindau-Webb, M. (1980). The story of /r/. UCLA Working Papers in Phonetics 51, 114-119.

Lindau-Webb, M. (1982). Phonetic differences in glottalic consonants. UCLA Working Papers in Phonetics 54, 66-77.

Maddieson, I. (1980) UPSID: UCLA Phonological Segment Inventory Database UCLA Working papers in phonetics, 50.

Nartey, J. (1982). On Fricative Phones and Phonemes UCLA Working papers in phonetics, 55, 1-181.

Nihalani, P. (1974). Lingual articulation of stops in Sindhi. Phonetica 30, 197-212.

Ramasubramanian, N. & Thosar, R.B. (1971). Synthesis by rule of some retroflex speech sounds. Language and Speech 14, 65-85.

Stevens, K.N. & Blumstein, S.E. (1975). Quantal aspects of consonant production and perception: a study of retroflex consonants. Journal of Phonetics 3,215-233.

Stevens, K.N. (1972). The quantal nature of speech: evidence from articulatory-acoustic data. Human Communication, A Unified View (ed. by P.B. Dennes and E.E. David Jr.) New York: McGraw-Hill.

Svarný, O. and Zvelebil, K. (1955). Some remarks on the articulation of the "cerebral" consonants in Indian languages, especially in Tamil. Archiv Orientalni 23,374-407.

Walldén, Ruth (1974-75). Contributions to the study of the pronunciation of i- and e- sounds in Tamil, especially before retroflex consonants. Orientalia Suecana 23-24, 200-9.

Wierzchowska, B. (1965). Wymowa Polska. Warsaw. Panstwowe Zaklady Wydawnickw Szkolnych. [not seen; as cited in Stevens & Blumstein 1975].

Wittgenstein, L. (1958). The Blue and Brown Books. Preliminary studies for the philosophical investigatons. New york: Harper and Row.

Phonetic and Phonological Representation of Stop Consonant Voicing

Patricia A. Keating

In this paper I argue for a more structured view of the relation between the phonological feature [voice] and its specific phonetic implementations. Under the theory of universal phonetics proposed here, the implementation of [voice] is sharply constrained: the opposition is defined relatively, as more or less voicing, along a dimension consisting of exactly three discrete, ordered categories, which can be shown to have a clear articulatory and acoustic basis. The phonological feature allows certain rule equivalences across languages to be expressed, while the phonetic categories describe possible contrasts within languages and express markedness relations.

1. Introduction

It is common practice to represent sounds in different languages which are phonetically similar but not absolutely identical with the same symbol or set of feature specifications. In some cases this practice is simply a convenience, for the sake of ease in transcription or typesetting. In other cases, though, the sounds in question are said to be different only at the phonetic level; at the phonemic level they may indeed have the same feature specifications, and so are thought of as the 'same' sound. The distinction in the _SPE_ model (Chomsky and Halle 1968) between systematic phonemic and systematic phonetic levels, for example, makes such a view possible. However, in this paper I will argue that the version of the model proposed in _SPE_ can be improved, allowing a more adequate treatment of a variety of facts. I will discuss how surface phonetic variation within and across languages can be derived in a synchronic grammar from the interaction of three relatively simple systems: the possible phonological features and their values, their possible phonetic category mappings, and phonetic detail rules accounting for variation within these phonetic categories. More generally, this paper contributes to an important goal of linguistic phonetics, relating discrete and timeless phonological units to physical reality, with its continuous articulatory and acoustic manifestations.

1.1. The _SPE_ model

In the phonetic theory of _SPE_, lexical items are represented as matrices of binary-valued phonetic features, with each row being a feature and each column being a segment. Phonological rules may change the values of features, or may add or delete segments, but may not change the inventory of features which form the rows of the matrices. Other, phonetic, rules convert the binary values into quantitative values along continuous phonetic scales; the rules depend on the value along each scale at which the given language divides the scale into its phonetic categories. Languages differ phonetically in part in just this way. These rules, like the binary phonological rules, are part of the 'phonological' component of the grammar, and no claim is made that quantitative rules are necessarily ordered after the binary rules.

The output of this phonological component is the phonetic transcription, containing all facts about pronunciation that are determined by the grammar. The phonetic scales available as rows of the phonetic matrix are universally fixed, but the quantitative values found along those scales will differ across languages. A further universal phonetic component, not technically part of the grammar, will convert these scalar values into a representation of articulations that are continuous in time. Given the universal phonetic scales, this translation or realization is assumed to be quite automatic. Various universal phonetic principles will produce phonetic patterns that will not need to be specified by the grammar of any one language.

Phonological feature representation serves to describe natural classes in rule application and to differentiate underlying forms. Phonetic feature (scalar) representation provides more detail, serving to describe any possible systematic differences between sounds that could distinguish two languages and might therefore be supposed to be part of the grammar of any one language. Following Ladefoged (1978), we might suppose that the appropriate level of detail to be represented is that necessary to distinguish a native from a foreign accent. Note however that phonetic feature representation is still removed from the physical level. In this model, it is the last form of discrete segmental representation prior to a continuous-in-time physical representation, the exact form of which remains to be established.

1.2. The case of consonant voicing

In this paper I will explore how this framework can be used to describe variation in stop consonant voicing within and across languages which have two phonemic classes ('voiced' and 'voiceless'). The discussion will be limited to this one case for two reasons. First, as will be seen, a wide variety of data will be brought to bear on the issues, and these data are available mainly for voicing. Second, there has traditionally been confusion over the terms 'voiced' and 'voiceless', and the phonetic symbols defined in terms of them, and they typically are used to mean different things in different languages. That is, in the case of voicing the problem is clearly defined already, and the relevant data are at hand.

The problem is this: the symbols for voiced and voiceless stops <u>b d g</u> and <u>p t k</u> are used for a variety of physical events. For example, Polish, like many languages, has a rather straightforward voicing contrast (to be discussed in Section 3) in which [b d g] are fully voiced and [p t k] are voiceless and unaspirated. English, on the other hand, has a great deal of positional variation, with its [b d g] being sometimes voiced and sometimes not (i.e. narrowly transcribed [b̥ d̥ g̊]) , and its [p t k] being sometimes aspirated and sometimes not. Common practice (e.g. Heffner 1969) reconciles the differences by describing both languages as having a voicing contrast, but English as having various allophonic rules which introduce, among other things, aspiration in certain environments.

However, in all of the recent rigorous attempts to define phonological features in phonetic terms, this common practice is abandoned. In the Jakobson, Fant, and Halle (1952) system, Polish would be described as having a voicing contrast, but English would not: it would have distinctive tenseness. Neither Chomsky and Halle (1968) nor Halle and Stevens (1971) even has a 'voicing' feature in the usual sense, and in both Polish and English would differ in the phonological features distinguishing /p t k/ and /b d g/. This is because these feature systems use physical features describing specific articulatory states to represent phonetic categories, as the basis for phonological representations. Recall that differences in scalar values express cross-language phonetic differences in the SPE system. Thus if Polish and English were to differ simply in the amount of aspiration required for a stop to count as voiceless, for example, they could be described as having the same phonological contrast despite their phonetic differences. But if they use different phonetic categories, or different articulatory mechanisms, then their phonological descriptions will necessarily be different, since it is the phonetic categories that are used as the phonological categories.

Consider how the features in Halle and Stevens (1971) describe voiced and voiceless stops of different kinds. They propose four features to characterize the state of the larynx at the moment of stop release: [±stiff vocal cords], [±slack vocal cords], [±spread vocal cords], and [±constricted vocal cords]. A combination of [-spread] and [+constricted] means a glottalized stop, of [+spread] and [-constricted] an aspirated stop, and of [-spread] and [-constricted] a plain stop. A combination of [+stiff] and [-slack] means a voiceless stop, of [-stiff] and [+slack] a voiced stop. If the vocal cords are simultaneously neither stiff nor slack, the segment type depends on the values of the other two features: a 'plain' stop is the voiceless lax stop found for many instances of English initial /b d g/; an 'aspirated' stop is the Korean partially aspirated type; a 'glottalized' stop is a true implosive. These features were originally designed to relate pitch to voicing differences with a single laryngeal mechanism; they have been criticized on that basis by, e.g., Anderson (1978).

The features are descriptively inadequate for voicing independent of its relation to pitch. For example, the difference between initial 'voiceless lax' (as in some English speakers' initial /b d g/) and initial prevoiced stops (as for other English speakers) is posited by Halle and Stevens to be one of degree of vocal cord slackness. In fact, the glottal configuration and state can be identical for these two categories. The determining factor for closure voicing is the pressure in the oral cavity: if oral pressure is kept low, either through cavity expansion or leak, closure voicing will occur (Westbury 1983). On the other hand, the difference between English medial voiceless unaspirated and (fully) voiced stops (e.g. rapid vs. rabid) is posited to be one of vocal cord stiffness by Halle and Stevens. In fact, the difference is more likely to be that the medial voiceless stops involve a glottal spreading gesture (Lisker, Abramson, Cooper, and Schvey 1969); by the time of release, the vocal cords will be back together and may well be slack just as for the voiced stops. Thus the glottal state at the moment of consonant release (the basis of the Halle and Stevens features) can be identical for English initial 'voiceless lax' and initial prevoiced, as well as for medial 'voiceless unaspirated' and medial voiced-through stops, requiring the addition of features for supraglottal pressure and occurrence of gestures at times other than release.

While such feature additions may save the system from these particular criticisms, this would be the wrong direction to take in enumerating phonetic features. A set of features predicated on phonetic accuracy will require ever more additional features as new articulatory mechanisms are discovered. The proliferation of features is the price paid for using the same set of features for phonological as well as low-level phonetic representation, an otherwise appealing constraint on the relation of phonology to phonetics. But at the same time, no claim is made that these largely redundant features could structure linguistic contrasts. For example, the Halle and Stevens features distinguish voiced laryngealized stops from true implosives, whereas languages never do (Greenberg, 1970; Ladefoged, 1982); similarly, they distinguish voiceless unaspirated from voiceless lax stops. The point, then, is that Halle and Stevens (and Chomsky and Halle) don't simply have the wrong features in these instances; they will _always_ have _too many_ features because they want to describe exactly how individual sounds are articulated. While we want the phonological features to have some phonetic basis, we also want to distinguish possible contrasts from possible differences. Our goal must be to find some feature framework in which the phonetic basis of phonological features is not explicit phonetic details.

1.3. A direction

It's clear that in the _SPE_ type of feature system, Polish and English /b d g p t k/ could look very little alike, and could not be referred to by the same symbols in a phonological representation. My starting point here will be to consider how the alternative result might be allowed, by relaxing the constraint that phonological feature representations incorporate specific articulatory information. Instead of concentrating on accurate phonetic descriptions of individual segments, let us consider how each level of representation could characterize some aspect of sound systems.

Suppose we try to modify the _SPE_ features so that, within the general model of the relation between phonology and phonetics, we have only as many phonetic categories given by the phonetic features as there are contrasting phonetic types in languages. Let us call these the _(major) phonetic categories_ and think of them as defining the phonetic symbols we normally use in transcription. That is, the categories are imposed along the phonetic scales underlying the phonological feature contrasts. Smaller-scale possible differences will not be represented in this transcription, as in fact they typically are not. Suppose further that the phonological representation will not directly reflect phonetic facts, but rather will organize natural classes for phonological rules. A phonological feature, then, corresponds to the use of terms like 'voicing' for various kinds of contrasts. Will such revisions in the feature system do any more than justify the traditional, nonrigorous, use of terminology? What could it mean to say that two languages have the same contrast, like voicing, when the corresponding physical events are somewhat different?

2.1. A proposal

Here we will explore the consequences of a proposal by Lieberman (1970, 1977): that [±voice] be used as a binary phonological feature which can be implemented differently in different languages along the continuous dimension of Voice Onset Time (VOT). VOT was originally proposed by Lisker and Abramson (1964) as the time between the onset of voicing and the release of a stop consonant; when voicing onset follows the release by any appreciable amount of time, aspiration occurs during that time interval. Thus the VOT dimension relates aspiration to voicing phonetically. However, Lisker and Abramson (1964, 1971) provided no suggestion concerning the role of this phonetic dimension in phonologies or in phonological feature representation.

As given by Lieberman, this proposal simply amounts to a claim within the _SPE_ framework that the physical scale appropriate to a voicing feature is the _VOT_ scale, and that plus and minus values of the voicing feature will have different quantitative VOT values in different languages. However, this specification does not deal with the question of the possible number of contrasts available to languages, since presumably any arbitrary set of VOT values could be specified under Lieberman's proposal. Therefore we will modify the proposal so that the binary phonological feature values will be implemented as VOT categories chosen from a fixed and universally specified set. This set, which will be motivated below, consists of (phonetically fully) voiced, voiceless unaspirated, and voiceless aspirated stop consonant categories. While they correspond directly to the standard division of the VOT continuum into lead, short lag, and long lag values for stops in initial position, they should be viewed as more abstract categories that include a number of acoustic correlates and articulatory mechanisms. To keep the phonological representation distinct from the phonetic representation, the convention will be followed that [voice], [+voice], and [-voice] refer only to the classificatory feature and its values, while {voiced}, {voiceless unaspirated}, and {voiceless aspirated} refer to the major phonetic categories. As in the SPE model, these categories will be further realized as articulatory and acoustic parameters represented continuously in time. To some extent these mappings will be part of the definition of the phonetic categories and therefore universal (e.g. {voiced} will involve vocal cord vibration and low frequency periodicity during consonant closure), but to some extent they will be language specific, in ways to be discussed below.

Given that the phonetic categories encode possible contrasts, we can return to the question of English vs. Polish /b d g p t k/. Both languages may be said to contrast [+voice] and [-voice] stops, but the phonetic categories which implement the phonological contrast will differ. For example, in some cases the Polish [+voice] stops will be {voiced} while the English ones will be {voiceless unaspirated}, and the Polish [-voice] stops will be {voiceless unaspirated} while the English ones will be {voiceless aspirated}. The details of these differences will be the topic of Section 3. For now, it suffices to note that the framework allows us to say that the two languages' stops are always the same phonologically though they may be different phonetically.

To summarize, three kinds of representation are being proposed. One is phonological: just as many features and feature values as are needed to distinguish the natural classes in a given language. Another is major phonetic categories, or modified systematic phonetic: just as many phonetic categories as are needed to distinguish possible contrasting categories in any language. The last is pseudo-physical: continuous in time and encompassing as many parameters as necessary for phonetic description; possibly, for example, autosegmental in character (Goldsmith 1976).

Before the implications of this proposal can be discussed, it should be made explicit that a limited set of cases are being considered. Basically we are dealing only with languages which contrast no more than two phonetic categories in any one context. It will be assumed, lacking evidence to the contrary, that all such contrasts should be described as [±voice]. We will exclude languages like Thai and Hindi, with more than two contrastive categories, largely because it is unclear whether such languages should be analyzed as having one feature [voice] which is not binary, or more than one binary feature. While it is immaterial to the arguments to be advanced here how these cases are treated, their analysis would certainly be relevant to a more complete account of voicing contrasts and their possible phonetic implementations. However, such analyses are beyond the scope of this paper.

More than a minimum of data would be required to bear on this question. Consider, for example, Thai, which in initial position contrasts voiced, voiceless unaspirated, and voiceless aspirated stops at all but its velar place of articulation (Haas, 1956). In final position only unreleased, largely voiceless, stops are allowed (Abramson, 1972). At first glance, this limitation appears to be phonological evidence that these are three values of a single ternary feature, since we hesitate to say that two independent features happen to neutralize (in the sense of defective distribution) in the same environment. But when we take further data into account, we can see that more than voicing is involved, since in final position the only coronal obstruent found is [t^{c}]. Thus not only voicing, but also place and manner features, are limited in final position, and the evidence provided for the unitary nature of a phonological [voice] feature in Thai is at best equivocal.

2.2. The phonological feature

Limiting ourselves, then, to cases like Polish and English, what does it mean to say that there is a level of representation in the grammar at which, for example, various phonetic sorts of /b d g/ are all [+voice]? What kind of evidence could be relevant in supporting this claim? I want to argue that evidence can be found in a prediction that the system makes about the relation between phonological rules and phonetic categories. If a rule applies to binary values, it cannot know what phonetic values those binary values are mapped onto. That is, they refer only to plus and minus values of the classificatory features in checking whether to apply to a particular form. This must be the case in this system, as long as the binary rules apply before phonetic category implementation, since rules cannot look ahead to future steps in a derivation. A rule which changes a binary feature value cannot know whether, for example, a resulting [+voice] value will be implemented as {voiced} or {voiceless unaspirated}. A curious consequence of this lack of look-ahead is that it should be possible for such a rule to occur with <u>either</u> sort of phonetic category implementation. That is, the occurrence of a phonological rule in a language should not depend on, or be correlated with, the phonetic details of the language. One could well object to this seemingly implausible prediction, since we tend to think of the occurrence of a rule as being intimately tied to its phonetic effects. Note that this issue does not even arise in a system where phonetic category differences are captured by different features, as in <u>SPE</u>.

However, this prediction is born out by various data. It appears that a distinction between phonological and phonetic category levels of representation offers an important advantage in describing phonological rules. In a system like <u>SPE</u> or Halle and Stevens (1971), which equate phonological with phonetic representation, various rules that occur across languages will look different in each language, depending on the phonetics. In a system like the one proposed here, which distinguishes the two levels, these rules will look the same regardless of the phonetics. Thus if there are rules affecting voicing that recur consistently across languages that differ in their phonetic categories, we will have evidence in favor of distinguishing phonological from phonetic representation. The generalization that certain rules occur across languages will be missed if phonological rules apply to phonetic features which are different across languages, but will be expressed if phonological rules apply to phonological features that are more similar across languages. I will present three relevant cases of rules referring to 'voicing' in a similar way across phonetically different languages, supporting the existence of a [voice] feature.

For example, Chen (1970) considers vowel duration before word-final 'voiced' and 'voiceless' stops in several languages and shows that in all of these languages, vowels are longer before 'voiced' stops than before 'voiceless' stops.[1] This apparent generalization would have to be given up if phonetic differences were incorporated into phonological feature representation as required in standard theories, since some of these languages would be described by various phoneticians as having 'tense' and 'lax' stops, or whatever feature is used, but others 'voiced' and 'voiceless'. That is, while vowels are longer before phonetically 'voiced' stops, and before phonetically 'lax' stops, we can conclude that in general they are longer before phonologically [+voice] stops. Further supporting data can be found in Mack 1982, which used precisely matched word lists for English vs. French /b d g p t k/ word finally in monosyllables. French, which contrasts {voiced} stops (with fully voiced closures) with {voiceless unaspirated} stops could be described as having a traditional 'voicing' contrast, unlike English. Mack found that in English, vowels before [-voice] stops were 53% as long as vowels before [+voice] stops, while in French, the ratio was 74%. Both of the English sets of vowels were longer than the corresponding French sets (154 and 285 msec vs. 133 and 182 msec for English vs. French).

Given these results, it could still be thought that vowel duration is phonetically governed, with, let us say, vowel duration being proportional to the phonetic 'voicedness' of the following stop - with the longest vowels before the most voiced stops, and the shortest vowels before the most aspirated stops. Two arguments can be made against this account. First, Mack's data show that the prediction is strictly false. Presumably the ranking for degree of phonetic voicing in her data is (from least voicing to most) English [-voice] < French [-voice] < English [+voice] < French [+voice], yet this does not correspond at all to the vowel duration ranking, French [-voice] < English [-voice] < French [+voice] < English [+voice]. Degree of lengthening does not appear to depend on degree of phonetic voicing: witness English, which has less closure voicing in its 'voiced' final stops than many other languages, given its tendency to partially devoice final [+voice] stops, yet more vowel lengthening. Hyman (1975, p. 171) actually describes English as having more vowel lengthening <u>because</u> it has less consonant voicing.

Second, consider languages with three or four categories: does vowel duration line up with degree of voicing across those categories? Maddieson (1977) looked at vowel durations before voiced and voiceless aspirated and unaspirated stops in Assamese, Bengali, Hindi, Marathi, and E. Armenian. The predominant, but not uniform, result was for vowels to be longer before voiced than before voiceless stops, and longer before aspirated than before unaspirated stops. That is, taking labials as an example, the effect of consonants on vowel duration could be ranked as:

$$p < p^h < b < b^h$$

with the effect of voicing being pairwise larger than that of aspiration. Presumably because Maddieson used a carrier phrase, but possibly due to measurement criteria, all of his duration values are much smaller than most of Mack's, so direct comparison between his three and four category languages and her two category languages is difficult. But once again we see that {voiceless aspirated} stops, the least phonetically voiced of all, do not cooccur with the shortest vowels. These examples have been given, despite our lack of a phonological analysis for them, simply to show that the determinant of vowel duration differences cannot be the position of the phonetic categories along the phonetic voicing dimension.

Another suggestion that vowel length does not depend mechanically on phonetic detail comes from the case of English vowels before flaps. Phonetically, flaps may be voiced or voiceless, but Fox and Terbeek (1977) showed that the duration of a preceding vowel, for speakers who make a distinction at all, depends on the underlying phonological value of [voice], and is not correlated with surface phonetic voicing. These types of data all argue that vowel duration is not conditioned directly by the degree of voicing during a stop consonant; whatever phonetic correlations may be found, a more compelling correlation is that between vowel duration and the value of the phonological feature [±voice]. This correlation can be expressed only if there is a phonological feature that is independent of phonetic categories.

Cluster voicing assimilation is another common phonological rule which appears to apply generally across phonetic categories. For example, Polish has regressive voicing assimilation (Mikoś 1977) and a [voice] contrast of {voiced} vs. {voiceless unaspirated} stops, while Danish has progressive 'voicing' assimilation but an aspiration contrast in initial position (Fischer-Jørgensen 1954).

Similarly, some evidence on fundamental frequency after stop consonant release indicates that phonological [voice] values are more important than phonetic voicing and vocal cord configurations in determining pitch patterns (Hombert, Ohala, and Ewan 1979). Citing data of Hombert and Ladefoged, Hombert et al. note that English {voiceless aspirated} [-voice] stops and French {voiceless unaspirated} [-voice] stops perturb the f0 of a following vowel by about the same magnitude. More recent work by M. Caisse (in progress, at Berkeley) makes a similar point. {Voiceless unaspirated} stops in initial position differ in f0 across languages depending on whether they are [+voice] or [-voice] (my paraphrase). The f0 differences must, of course, have some articulatory cause, and in that sense there must be two kinds of voiceless unaspirated stops -- Hombert et al. suggest a difference in larynx height. However, their distribution appears to depend on their phonological function (what they contrast with). The pitch perturbation rule is most simply stated in terms of phonological feature values.

Thus the distinction between phonological and phonetic features appears to be not only plausible, but necessary if rules such as those discussed are to be properly defined across languages. Other rules that come to mind as depending on phonetic content of segments can best be described under the phonetic implementation rules, for example, the fact that certain phonetic segments alternate as implementations of a [+voice] specification. Statements about segment frequencies in phonological inventories (e.g. lack of /g/ relative to /k/) may also depend on segment content. In Section 4 below it will be seen that one function of the phonetic level of representation is precisely to allow for this kind of statement about markedness, outside of the set of synchronic rules contained in the grammar. While it is conceivable that rules can be found which must refer to the phonetic implementation of a phonological segment or contrast, the fact remains that for at least some rules, the proposed phonological representation elucidates cross-language generalizations.

2.3. Phonetic Categories

Having seen that somewhat phonetically abstract phonological feature values are necessary, we turn now to motivating the phonetic categories that implement these values. A careful consideration is required because positing such categories along the phonetic voicing dimension is one of the main modifications of the SPE framework being proposed here. These categories will be defined along a phonetic voicing dimension that in initial position coincides with the VOT dimension. Four steps are necessary. First, we must define VOT and the voicing dimension. Second, we must show that languages are limited to three contrasting categories, that is, that three is the right number of major phonetic categories to posit. Third, we want to see that the three categories are the same three in various languages, in particular, the three posited here as being provided by the phonetic theory. Fourth, we will examine the functional, extralinguistic, basis of these categories.

We begin with the description of voicing and VOT. As the time interval between the release of a stop consonant occlusion and the onset of vocal fold vibration, VOT is meant to be a cover term for various laryngeal and supralaryngeal events associated with this timing relation. The acoustic manifestations of the so-called VOT dimension are diverse. In practice, VOT is measured from acoustic displays as the time between the release burst and the first quasi-periodicity in the acoustic signal, and this time interval is referred to in a narrow sense as acoustic VOT. VOT is usually represented graphically as a continuum of time values. Stop release is the arbitrary reference point in time, 0 msec. Voice Onset is measured relative to that point. Voice Onset occurring coincident with stop release is called 0 msec VOT. Voice Onset occurring before stop release is assigned a negative VOT value and is said to lead the release. Voice Onset occurring after stop release is assigned a positive VOT value and is said to lag the release. Positive VOT values to about 20-35 msec (depending on the place of articulation) are called 'short lag', and higher values are called 'long lag'. In general, stops traditionally described as being voiceless unaspirated (or devoiced) have short lag VOT's, and stops traditionally described as being voiceless aspirated have long lag VOT's.[2]

What then of other positions in the word, and their relation to the VOT dimension? Recall that {voiced} stops in initial position are characterized by voicing lead, that is, voicing during stop closure. This voicing may or may not begin at the moment of closure, and may or may not extend throughout the entire closure, but there is at least some glottal vibration and low-frequency periodicity during closure. In other positions as well, {voiced} stops are characterized by voicing during stop closure. If the stop follows a sonorant, then the voicing of the sonorant and the stop closure will typically be continuous. No measurement can be made of the time of voice onset, since voicing is already on. However, a measurement of closure duration, from the onset of this voiced closure to stop release, is most equivalent to a VOT measurement for {voiced} stops following pause. If a {voiced} stop is in final position, or before another stop, it may not be released, in which case the only possible measurement would be the amount of voicing during the closure. In all of these cases, regardless of which endpoints can be ascertained, the stop closure crucially contains some low-frequency vibration.

Similarly, in all positions a {voiceless aspirated} stop has a measurable amount of aspiration after the release. In medial positions before a sonorant the measurement of VOT proceeds just as for initial position. While there may be one or two pitch periods of voicing at the beginning of closure, carrying over from a preceding voiced segment, closure voicing does not extend beyond that small amount. The only difficulty arises when the {voiceless aspirated} stop is not released into a sonorant, and so no VOT measurement is possible (since voicing does not begin again). In these cases the duration of the aspiration, which is quite audible, can often be measured from acoustic displays. Closure duration for {voiceless aspirated} stops is typically fairly short.

A {voiceless unaspirated} stop is one with no more than a few pitch periods of voicing during its closure, and a short lag interval of voicelessness following its release (if any). This category differs from {voiced} in its relative lack of closure voicing, and from {voiceless aspirated} in its short lag VOT values following release. If there is no release, the two {voiceless} categories cannot be distinguished. Data from Polish listeners (Keating 1979) suggest that a stop closure must be about one-half voiced for a Polish medial /d/ to be heard.[3] While such phonological judgments do not, of course, directly indicate perception of the phonetic categories, they are relevant to the observation that in final position degrees of closure voicing do not appear to be distinguished among various stops classified as [-voice], although they logically could be. We have already seen that fundamental frequencies after {voiceless unaspirated} stops vary; closure durations also are not uniform. In sum, this category appears to show the most acoustic variation of the three, except that when VOT can be measured, it is confined to the narrow short lag region of the VOT continuum.

Now we can return to the question of how languages use the phonetic dimension just described for stop contrasts in word initial position. Lisker and Abramson (1964) pointed out that no language appears to contrast more than three categories along the VOT dimension. If more than three categories are contrasted, at least two of them will have similar VOT values, and they will differ along some other dimension. Certainly a survey of the traditional literature on a number of languages supports the view that three general categories are sufficient for descriptions of contrasts and even for most cases of allophonic variation. All of the 51 languages surveyed by Keating, Linker, and Huffman (1983, this volume) used at least some kind of voiceless unaspirated stops in virtually every position, at least according to sources. As categories contrasting with them, fully voiced and voiceless aspirated stops are about equally common. Voiced aspirate, prenasalized, 'tense', and implosive categories are also found, but in these cases the VOT values are the same as for one of the three basic categories. Among the languages contrasting no more than two VOT categories in any one environment are English, German, Spanish, Polish, French, Tagalog, Dutch, Swedish, Mandarin, Cantonese, Russian, and Kirghiz. Languages using all three, at least in some environment, include Thai and E. Armenian. Languages using these three plus some other include Hindi and other languages of India, which add voiced aspirates. In sum, cross-language distributional evidence supports Lisker and Abramson's idealization of three basic voicing categories for contrastive and allophonic use in initial position. And in fact, they are enough elsewhere, since no greater number of contrasts is found in any other position.

Let us turn now to the evidence for three particular phonetic categories, and the boundaries between them. Lisker and Abramson observed that while there are differences across languages in exactly what VOT values are used, there is rough agreement on the areas of the VOT continuum. These areas correspond to the traditional phonetic categories of prevoiced, voiceless unaspirated, and voiceless aspirated stops, taken here to be components of the feature system.

Following this observation, it is generally assumed that there is some universal framework which provides these categories, and that the cross-language variability is secondary in importance. Let us consider this assumption in some detail.

If no language requires us to posit more than three categories along the phonetic voicing dimension, can we go further and say that the same three categories are used in different languages? It could be that each language divides up the continuum rather differently, even if always into three or fewer categories. While the identity of categories across languages is not crucial to the formal system being proposed, it would offer a further constraint on possible phonetic implementations and would be relevant to questions of the contribution of a universal phonetic theory to feature systems. Here we will review the evidence that there are really three discrete categories, rather than fuzzy areas of the continuum, found across a variety of languages. Next we will turn to possible functional explanations for the locations of these categories.

As a baseline on the division of the voicing dimension into three categories, let us refer to languages with three or more contrastive phonetic categories. Lisker and Abramson (1964) give data on VOT in initial position for a number of such languages. As an example, consider Thai, with three contrasting categories in initial and medial position for labials and apicals. The {voiced} stops have lead VOT values, up to about -40 msec; the {voiceless unaspirated} stops have short lag values, from 0 to +10, +20, or +30 msec VOT, depending on place of articulation; the {voiceless aspirated} stops have higher VOT values. Donald (1978) showed that Thai listeners have discrimination peaks at about -20 and +20 msec VOT, corresponding to these distributions.

In the Thai data, there is an obvious gap between the lead and short lag VOT values, that is, a lack of low negative values. This gap can be seen in data presented below in Section 3 as well, and in fact seems to be the general rule for languages with any number of voicing categories. Whatever its cause, it has the effect of clearly separating {voiced} from {voiceless unaspirated} stops acoustically, and makes the {voiced} category largely coincide across languages. Furthermore, the {voiceless unaspirated} and {voiceless aspirated} categories as they occur in Thai appear to be quite similar in other languages that contrast such categories. Inspection of the VOT values for {voiceless unaspirated} stops in a variety of languages indicates that they typically lie within a narrow area of the VOT continuum and are essentially normally distributed within that area. (This normal shape is more apparent in graphs plotted to a finer scale than 10 msec intervals, as is done here and in most spectrogram-based studies.) Although there are slight differences across languages in these distributions, in general we can say that labial {voiceless unaspirated} stops have VOT values up to about +20 msec, apicals up to about +30 msec, and velars up to about +40 msec. The distributions appear to be constrained on the low VOT side by the 0 value and the gap, and on the high VOT side by the long lag values they contrast with. Crucially, the division between short and long lag values is quite similar across languages, and the corresponding perceptual boundary is also similar across languages (for a summary of the literature, see Keating 1979).

What about languages contrasting two categories, {voiced} and {voiceless unaspirated}? Because of the gap in VOT values, we expect the {voiced} values to look like those of Thai, etc., but there is no reason that the {voiceless unaspirated} values should be constrained on the high VOT side, and we might expect to see an acoustic contrast of lead vs. general-lag VOT. Generally this does not happen; {voiceless unaspirated} values in such cases typically resemble those that are constrained by a {voiceless aspirated} category. In medial and final positions, {voiceless unaspirated} stops are less clearly distinguished from {voiced} stops, since the amount of voicing during closure varies continuously.

To the extent that there are three fixed phonetic categories whose values are constrained in this way, then the case of stop consonant voicing provides a counterexample to the ideal of 'maximal dispersion' (Liljencrants and Lindblom 1972), by which languages keep their contrasts maximally distinct in the phonetic space. On this hypothesis the most favored contrast should be {voiced} vs. {voiceless aspirated}, with extreme VOT values. Such a contrast is at best rare in languages (Flege 1979). The most common category across all environments and languages is the {voiceless unaspirated} category; it is near universal, both alone and in contrast with one or both of the other categories, and its acoustic values are highly constrained.

However, there are some cases of phonological categories having phonetic values beyond these observed category limits. [+voice] stops which are basically {voiced} may have a few {voiceless unaspirated} tokens, and [+voice] stops which are basically {voiceless unaspirated} may have some {voiced} tokens. And, for {voiceless unaspirated} stops, as in the Polish data to be presented below, there may be a tail of values into the higher VOT range from high vowel contexts, or from extra emphasis, or for no apparent reason other than spreading over the phonetic space, as Polish /k/ does. Still, such spreading need not be the case. English speakers do not generally begin prevoicing, or prevoice more, in more careful speech.[4] And Spanish speakers do not prevoice their /p t k/ even in contexts in which VOT is not the only contrastive dimension. Dent (1976) hypothesized that since in running speech most instances of Spanish /b d g/ are spirantized, while /p t k/ are stops, (i.e. the voicing contrast is confounded with a manner contrast), then the {voiceless unaspirated} category would be free to extend into the {voiced} category. However, her acoustic measurements of VOT for [-voice] stops in contrasting and noncontrasting contexts showed no difference. Thus languages may have phonetic spreading, but do not always do so.

The discussion of the three phonetic categories so far has been limited to the question of acoustic similarities and constraints across languages. There may also be articulatory similarities. While cross-language physiological data is limited, it appears that glottal gestures for {voiceless aspirated} stops, for example, are quite similar for English and Swedish (e.g. Lofqvist 1980; Yoshioka, Lofqvist, and Hirose 1981). And despite various acoustic differences, Weismer (1980) found a constant voiceless (open glottis) interval for [-voice] stops and fricatives, even across place of articulation. Word initial clusters of /s/ plus a [-voice] stop also appear to have the same constant opening (Yoshioka et al. 1981), accounting for the fact that such stops are {voiceless unaspirated}. The glottal gesture may be quite similar across segments, contexts, and languages because it is a ballistic movement, not controlled for duration or extent. The same gesture can be used for both aspirated and unaspirated stops as long as the timing of the consonant closure is manipulated. If it were, then [-voice] stops of either category would have a constant glottal definition; however, the references cited indicate that some {voiceless unaspirated} [-voice] stops have a very small, or no, glottal opening. Thus the goal of the speaker may be the state of no vibration as much as the state of an open glottis. That the goal can be in terms of vibration is suggested by English speakers who prevoice their [+voice] initial stops. These speakers use an extra articulatory gesture to achieve the same output, {voiced}, as is found in medial position, where {voiced} stops require no extra gestures (Westbury 1983). On the other hand, Hayes (1983) presents evidence from Russian voicing assimilation that speakers' goals can be glottal states rather than vibration. Overall, then, the issue of what is being controlled to achieve what goals remains problematic; for the present purposes we can only note that various possibilities exist, and that the phonetic categories must be defined in all of the relevant domains. It is quite likely that in some ways the three phonetic categories have a functional motivation in terms of articulatory mechanics and control.

There is also evidence for a perceptual basis for three fixed phonetic categories. Boundaries between these phonetic categories can be elicited from listeners in languages where they do not represent a linguistic contrast: the boundaries are present but cannot have been induced by phonological experience. Both Abramson and Lisker (1972) and Williams (1974) found extra discrimination peaks for Spanish labials at about +25 msec VOT, between {voiceless unaspirated} and {voiceless aspirated} categories, besides the (linguistic) peak at about -10 msec VOT. English listeners can also show such a non-linguistic peak, between {voiced} and {voiceless unaspirated} stops, given the right experimental procedures (Pisoni 1977, Carney, Widin, and Viemeister 1977). And Kikuyu listeners, for whom VOT differences correlate with a contrast between {voiceless unaspirated} and prenasalized apical stops, show two discrimination peaks for labial stops, at -15 and +20 msec VOT (Streeter 1976). Thus we see that not only do languages agree on roughly where the perceptual boundaries fall, but listeners whose linguistic experience does not include those boundaries also put them in the same place.

This fact suggests that there is some extra-phonological basis for the two category boundaries, that the auditory system imposes a discontinuity on the perception of VOT that is exploited by linguistic categorization. Evidence that the discontinuity is extra-phonetic as well comes from studies of animal perception of VOT. Perception of VOT has been studied in rhesus monkeys (Waters and Wilson 1976) and chinchillas (Kuhl and Miller 1975, 1978; Kuhl 1978). The methodology of the Kuhl and Miller studies with chinchillas allows a more direct comparison with adult studies, although only in the lag VOT region.[5] The result is that the animals showed the same boundaries as adult English speakers on the same stimuli, indicating that at least that boundary must be attributed to non-linguistic properties of the auditory system.

Further research has been directed at identifying such psychoacoustic factors, starting from a hypothesis by Hirsh (1959, 1975) about limits on the ability of the auditory system to resolve and sequence acoustic events separated in time. Suppose that voicing onset and stop release are two separate events that must be perceived in the right order for a voicing judgement to be made; if they are too close in time (say, 20 msec) they cannot be ordered accurately; they will be perceived as being simultaneous, and the percept will be {voiceless unaspirated}. Following a suggestion to this effect by Stevens and Klatt (1974), Miller et al. (1976) and Pisoni (1977) applied Hirsh's work to the perception of nonspeech stimuli designed to be analogous to VOT stimuli. In discrimination tasks, most listeners showed peaks at the analog values of about +20 and -20 msec VOT. Some debate has ensued as to whether these results are enough to account for all of VOT perception. Whether or not they are, perceptual discontinuities along the VOT (and possibly other relevant) dimensions indicate that there is some psycho-physical basis for the three phonetic categories. Whatever that basis may turn out to be, it would constitute a functional explanation for the consistency of the categories across languages. While the phonetic categories have a formal representation in the grammar, they also have this nonlinguistic perceptual motivation, plus whatever articulatory bases may be identified.

3. Applying the system to data

Given these three major phonetic categories, how are they used in languages as implementations of phonological feature values? Here we will consider phonological to phonetic category mapping across allophonic variation within each of three languages: Polish with a rather simple system, and then English and German with more complex systems (Section 3.1). We will then see that the descriptive framework motivated by these within-language considerations is sufficient for cross-language comparisons as well (Section 3.2). This will necessitate a fairly detailed examination of a certain amount of acoustic data.

3.1. Contextual allophones in three languages

First consider voicing in Polish. The [voice] contrast in Polish is an extremely straightforward one. Polish contrasts /b d g/ with /p t k/ in initial and medial positions. Word-final stops before pause are neutralized to [p t k], but before a sonorant-initial word this neutralization is optional. Before an obstruent-initial word, a final stop is subject to regressive voicing assimilation (Mikoś 1977).

Figure 1 shows acoustic measurements of VOT in postpausal initial position (from Keating, Mikoś, and Ganong 1981). A list of 42 disyllabic words beginning with all phonologically legal sequences of a stop consonant [b d g p t k] followed by a vowel [i], [e], [ẽ] [a], [ã], [o], [u], [ɨ] was read ten times each by five monolingual Polish speakers in Łódź, Poland. Palatalized allophones of /t/ and /d/ before [i] were not included. The VOT distributions for the voiceless stops show a normal distribution in the short lag region, but skewed with some long lag values due to high vowel contexts. (High vowels generally cause higher VOT values, since pressure in the oral cavity behind the constriction is vented more slowly.) Nonetheless, it is striking how little overlap there is across vowel contexts.

Figure 2 shows acoustic measurements (from Keating 1979) for medial poststress position for [t] and [d], including lag VOT values for items without closure voicing, and closure duration for items with closure voicing, since voicing continues largely uninterrupted throughout the V[d]V sequence. While medial [+voice] stops have somewhat shorter closures and therefore less measured voicing than the initial stops do, the medial [-voice] stops have VOT values which are very similar to the initial ones. In Polish, therefore, stops in initial position and stops in medial position are closely similar. [+Voice] stops have voicing during closure and sometimes through the burst. [-Voice] stops do not, with voicing always beginning after the burst. That is, Polish appears to be an uncomplicated, uncontroversial case of a [voice] contrast in which the surface [+voice] members are {(fully) voiced}, and the surface [-voice] members are {voiceless unaspirated}.

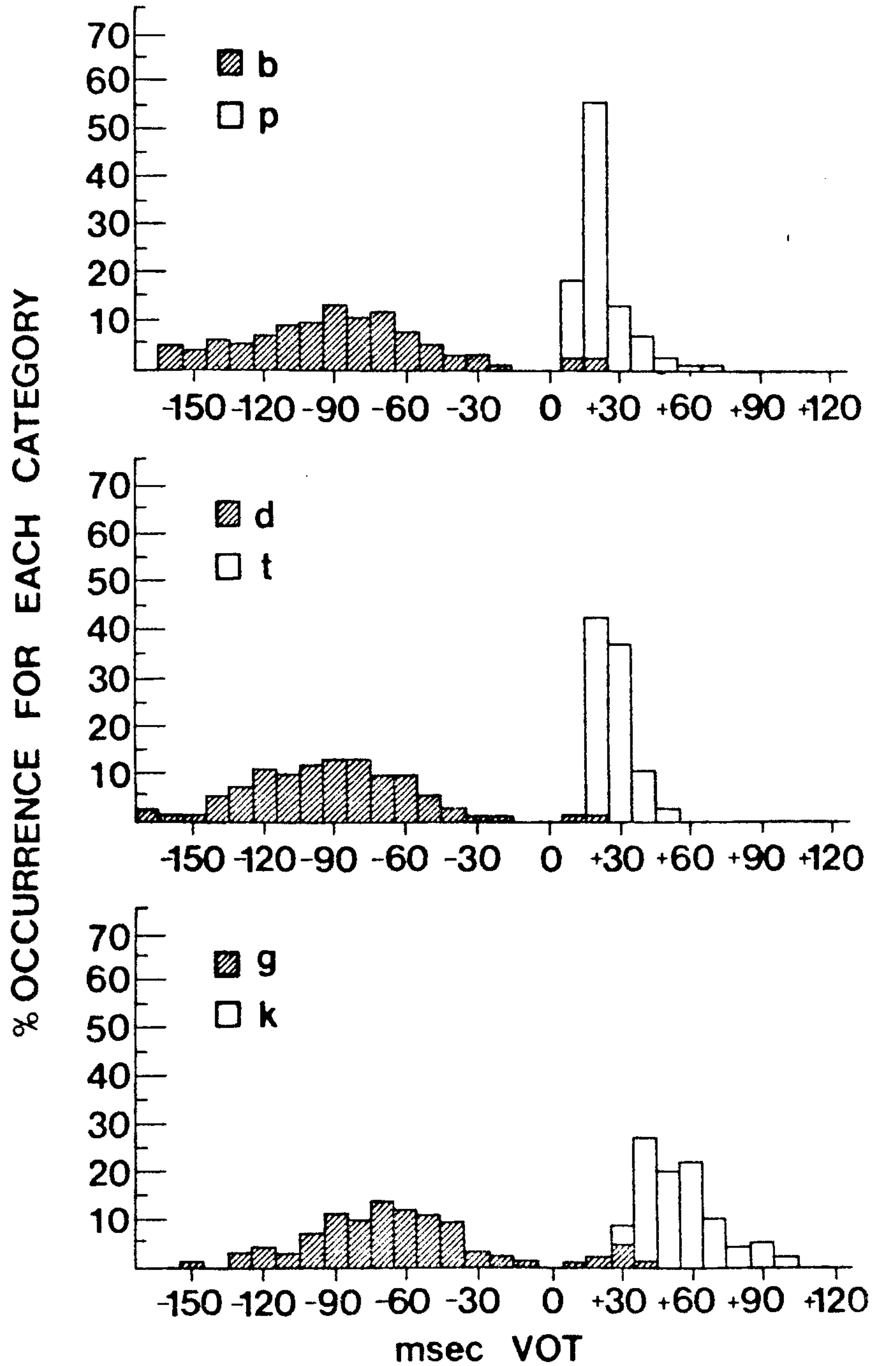


Figure 1. VOT values for Polish utterance-initial stops at three places of articulation before eight vowels. Measurements are for five speakers reading 42 words ten times each.

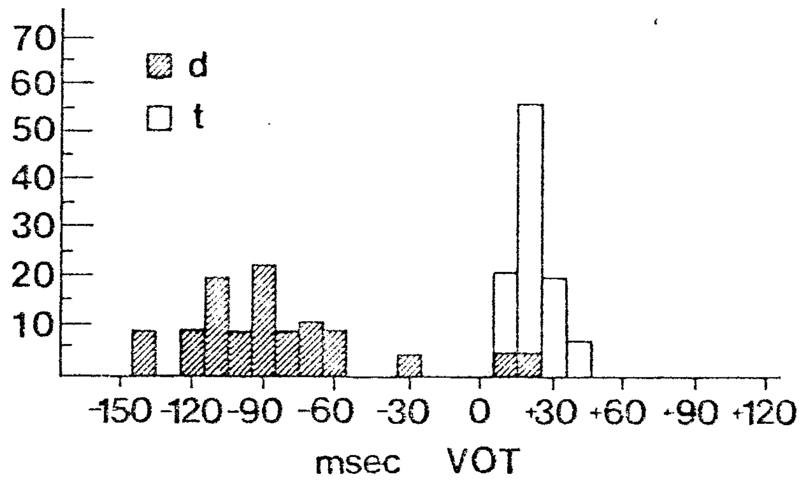


Figure 2. VOT values for Polish medial /t/ in <u>data</u>, <u>tata</u>, and <u>rata</u>, read once or more by 24 speakers, and voiced closure durations or lag VOT values for Polish /d/ in <u>rada</u>, read once or more by 22 speakers.

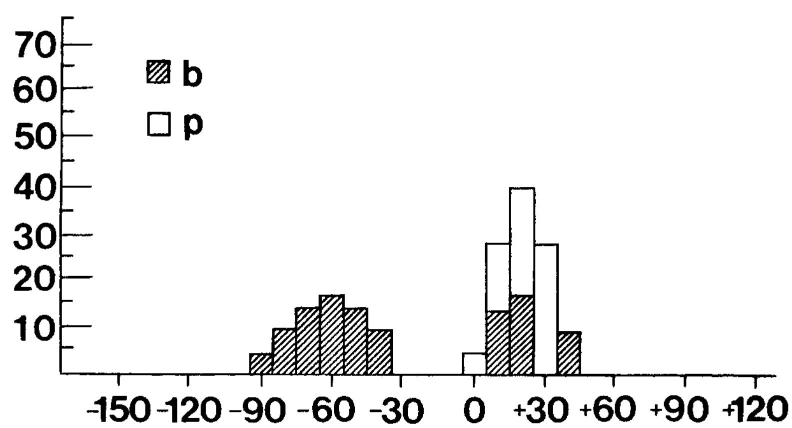


Figure 4. VOT values for English medial /p/, and voiced closure durations or lag VOT values for English medial /b/, from a list of words read once by six speakers.

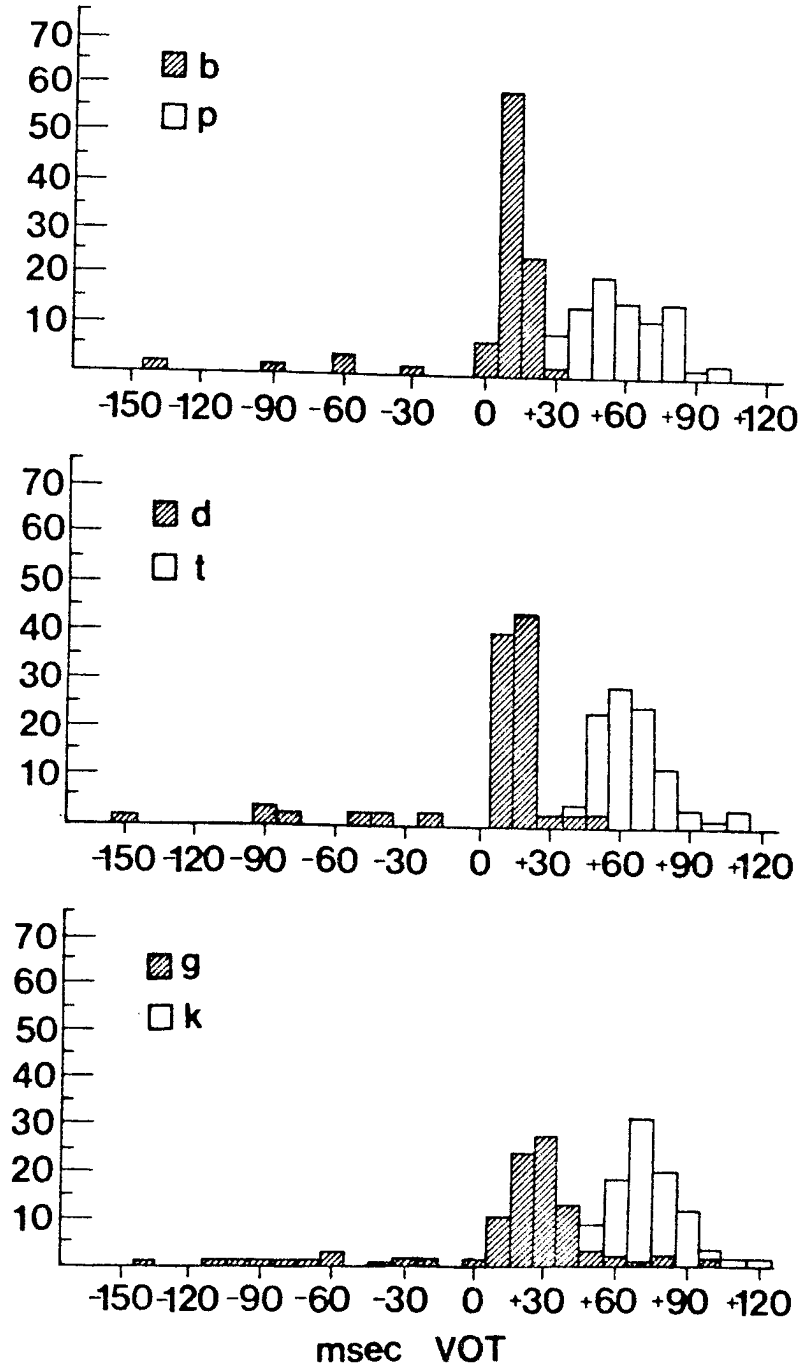


Figure 3. VOT values for English utterance-initial stops at three places of articulation before twelve vowels. Measurements are for two speakers reading 72 words, one four times each and the other two times each.

Compare Polish now with English phonetic implementation. It is well known that in initial position English [+voice] stops are {voiced} or {voiceless unaspirated}, and that the [-voice] stops are {voiceless aspirated}. This result was replicated in a way directly comparable to the Polish study described above. Figure 3 shows VOT measurements for English postpausal initial stops before 12 vowels, from a list of 79 disyllabic words read four times by one, and two times by a second, monolingual American English speaker (from Keating et al. 1981). It can be seen that English divides up the VOT continuum differently from Polish (Figure 1), with some lead values, but mainly short lag vs. long lag.

Next, compare these initial English values with medial poststress values, which have generally been noted to differ (Lisker and Abramson 1967, Flege and Brown 1982). Although the Polish data were for /t d/, in American English these are generally flapped in this context, so the English data presented are for /p b/. Six speakers read a list of words containing medial /p/ or /b/ before a reduced low vowel, resulting in 25 /p/ measurements and 24 /b/ measurements. The lag VOT values for /p/ are displayed in Figure 4 as the Polish data were shown in Figure 2. While all of the /b/'s had a substantial amount of closure voicing, the figure shows lag VOT values for stops which had any voicelessness during their closures, and closure durations for the others. The two languages differ in that the English voiced closure durations are shorter than the Polish. Overall, however, the values are remarkably similar in this context.

We see then that English has more variation in its phonetic implementation of [±voice] than Polish does. Such variation has been noted before; for example, Kahn (1976) contains a valuable study of the phonetic variants of English /t/, including [t^h], [t], [ɾ], and [$t^ʔ$]. General syllable-based rules for deriving these allophones are proposed, largely extendable to /p/ and /k/ as well. However, no corresponding analysis of voiced stops is given. It is often assumed in the phonological literature that only voiceless stops are subject to such rules -- that (ignoring flapping) /b d g/ are always 'voiced' while /p t k/ are voiceless but [±aspirated]. In fact, though, as we have already seen for one medial context, /b d g/ also vary phonetically, sometimes being voiced throughout their closures, and sometimes being voiceless unaspirated.

In addition to variation across contexts, English has also been shown to have substantial speaker differences. Previous studies (Lisker and Abramson 1964, 1967, Smith and Westbury 1975, Flege 1982) have shown that some English speakers produce some or all of their initial voiced stops with prevoicing rather than short lag VOT, although there are other conditioning factors involved.[6] Furthermore, Flege (1982) showed that speakers' glottal gestures, at least to the extent that they can be determined with an electroglottograph, differ in the production of English initial /b/. Some speakers open the glottis, while others do not. However, no single speaker had two glottal timing patterns for initial /b/. This was so even though individual speakers sometimes prevoiced and sometimes did not. That is, the observed glottal patterns were quite consistent across tokens while the occurrence of prevoicing was not. For a given speaker, the acoustic variation must have been due to additional articulatory mechanisms.

To summarize to this point, Polish implementation is quite simple: [+voice] as {voiced}, and [-voice] as {voiceless unaspirated}. English shows more variation, both across positions and speakers; [+voice] is implemented as {voiced} and {voiceless unaspirated}, [-voice] as {voiceless unaspirated} and {voiceless aspirated}. However, to be more sure of the variation that occurs in English, more data from a single group of speakers is required.

Six Americans, four female and two male, read a list of 215 words, each of which contained one of the six stops of English before a low vowel. Stops occurred either initially or intervocalically; the following vowel had primary stress, secondary stress (full vowel), or was reduced. (The medial unstressed /b p/ tokens are the ones already presented.) There were six words for all but one of the 36 combinations of stop, position, and stress. For each recorded token, up to three measurements were made from a computer-implemented oscillographic display. For postpausal stops, VOT was measured as described before. For postvocalic stops, the duration of voiced and voiceless closure intervals was measured, as well as lag VOT. The sum of the two closure measurements gives the total closure duration. Figure 5 shows a schematic waveform for each context for the labials only, where each section of the waveform represents the mean across the six speakers for that measurement.

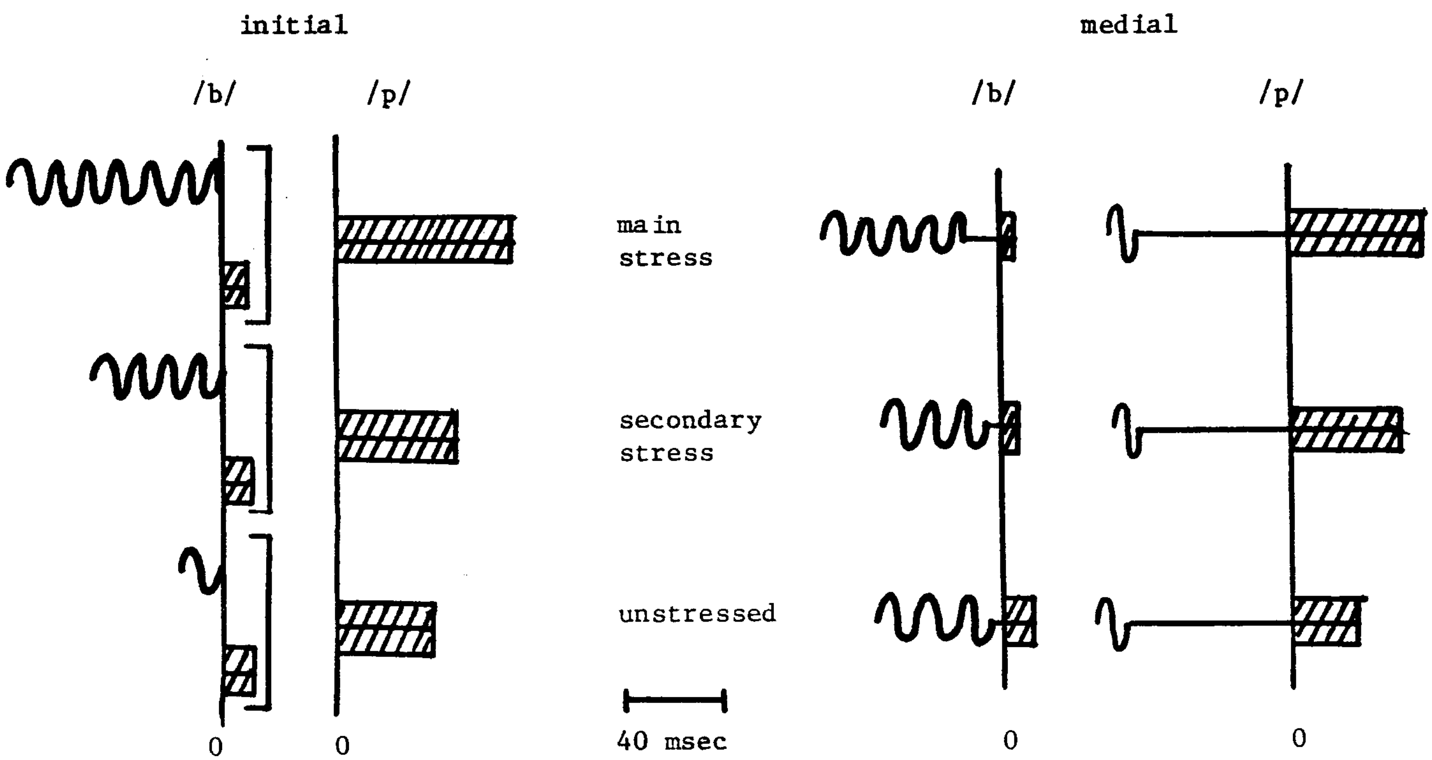


Figure 5. Phonetic implementation of English /b/ and /p/ in initial and medial positions, for three stress environments. The vertical lines, labeled '0', represent the moment of stop release. Each schematic shows mean values for the three measurements duration of closure voicing (wavy lines), duration of closure voicelessness (straight lines), and duration of voicing lag (striped bars). For initial /b/, prevoiced and short lag values are graphed separately.

One general pattern across the three places of articulation is that initial [-voice] stops usually have long lag VOT values, confirming Kahn's results, and initial [+voice] stops usually have short lag VOT values, regardless of stress. The degree of aspiration of the [-voice] stops is somewhat sensitive to stress level, although for initial /t/ the VOT values are the same before vowels with secondary stress and no stress.[7] The [+voice] stops are occasionally prevoiced; the occurrence and duration of prevoicing vary with place of articulation and stress. Another pattern is that medial [+voice] stops generally have voiced closures, although less voicing before main stress. Medial [-voice] stops have long lag VOT values before main and secondary stress. Before reduced vowels, the values are much lower, falling within the range of short lag, rather than long lag, values. Medial /t d/ before reduced vowels are of course typically flapped.

We see then that there really is fairly consistent stop consonant variation across contexts in English. Is this phonetic variation just an idiosyncracy of English, or is it more general? This pattern was investigated further by looking at a subset of these environments for one speaker of Standard German, another language with an initial aspiration contrast. This speaker read a set of words containing labial or alveolar stops before low and mid vowels: initially with main stress or with secondary stress, and medially before main stress or before a reduced vowel. Results of spectrographic analysis are summarized in Figure 6. Initial [-voice] stops are all somewhat aspirated. Medial [-voice] stops are somewhat less aspirated, but surprisingly, contrast with voiceless unaspirated [+voice] stops.[8] The exception to this generalization is medial /b/ before a stressless vowel, which is often voiced. That is, /b/ and /d/ differ with respect to medial voicing. Basically, however, the German speaker shows less variation in use of major categories, and more variation in the degree of voicelessness associated with the different environments, than the English speakers did. So it's not English alone that shows positional variation, but the variation in the two languages is not the same. Each language must have its own implementation rules.

From descriptions in the literature, it seems likely that such languages as Norwegian (Vanvik 1972), Mandarin (Dow 1972), and Kirghiz (Hebert and Poppe 1963), all languages with aspirated initial [-voice] stops, show some variation of this sort. For example, Kirghiz's [+voice] stops are described as being voiceless unaspirated initially, medially adjacent to another 'voiced' stop, and finally after a long or heavy-stressed vowel, but voiced elsewhere. The [-voice] stops are described as being voiceless aspirated initially, and voiceless unaspirated elsewhere. By this account, then, there is some neutralization in clusters and word finally after certain vowels, and otherwise two possible contrasts: an initial one of aspirated vs. unaspirated stops, and a medial and final one of unaspirated vs. voiced stops. While more thorough, acoustic, studies are needed of such cases, the general pattern appears to be that languages like Polish, without contrastive aspiration, show little allophonic variation, while languages like German and English, with contrastive aspiration after pause, show a great deal of positional variation. (See Keating, Linker, and Huffman, this volume.) If this is so, then the implementation rules of such aspiration languages will be somewhat complex, and they will apparently be at least partly language-specific.

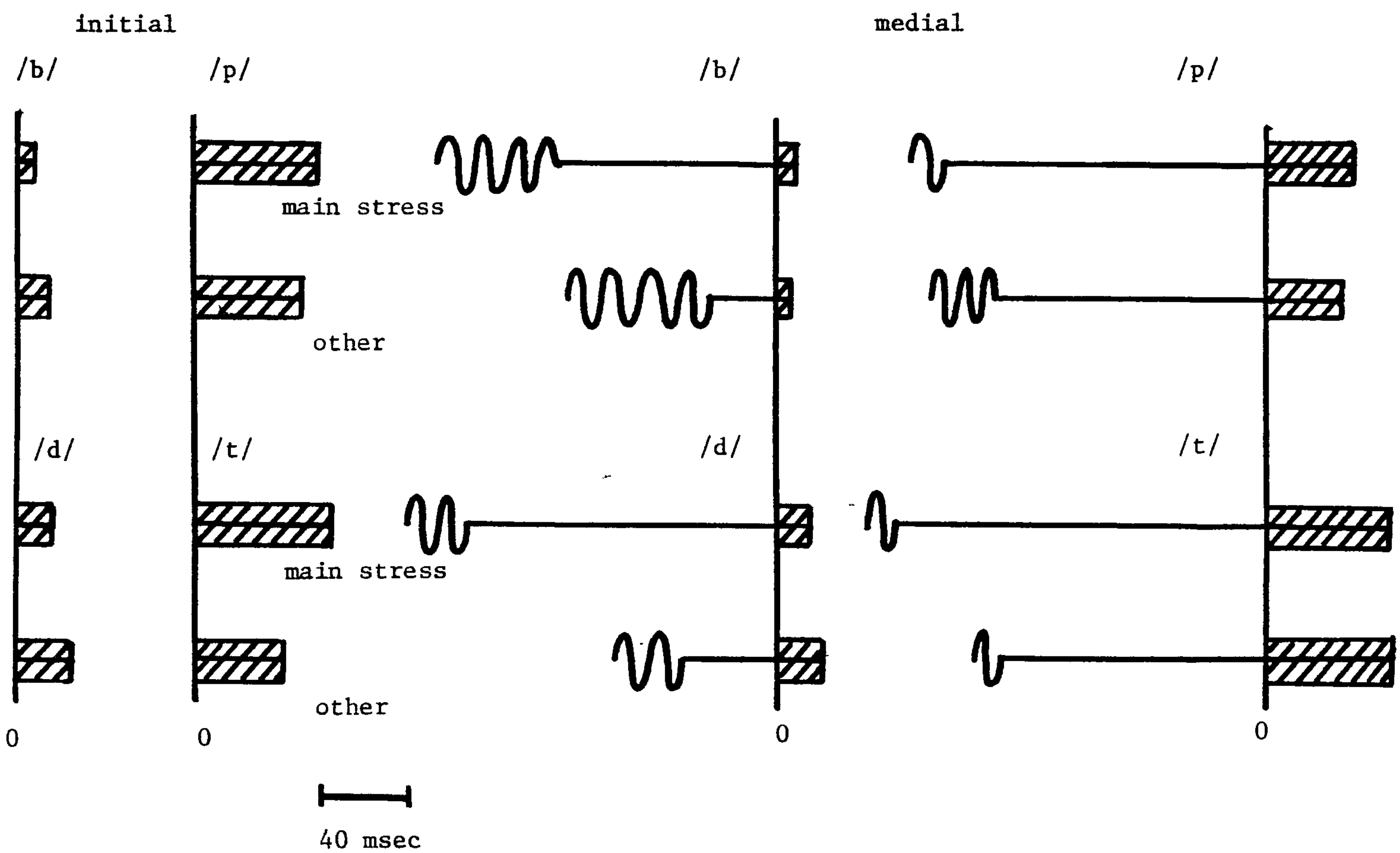


Figure 6. Phonetic implementation of German /b/, /p/, /d/, and /t/ in initial and medial positions for various stress conditions. The 'other' stress for initial stops is largely secondary stress, while the 'other' stress for medial stops is largely unstress. The form of the graph is like that in Figure 5.

3.2. Cross-language description

By considering quite a bit of acoustic data, we have seen three degrees of complication in the phonetic implementation of a [±voice] contrast. Polish, like other non-aspiration languages, is the least complex case: [+voice] is always {voiced}, and [-voice] is always {voiceless unaspirated}. English for speakers who prevoice, and German as spoken by at least one speaker, are somewhat more complex, since two implementations are used in each case. For this kind of English, [+voice] is {voiced}, and [-voice] varies between {voiceless unaspirated} and {voiceless aspirated}. For German, on the other hand, [+voice] varies between {voiced} and {voiceless unaspirated}, while [-voice] is {voiceless aspirated}. The most complex case is English for speakers who do not prevoice: [+voice] varies between {voiced} and {voiceless unaspirated}, and [-voice] varies between {voiceless unaspirated} and {voiceless aspirated}, sometimes in a correlated fashion.

While there is clearly no general pattern across languages, it is also true that phonetic implementation of [voice] is constrained somewhat. Not surprisingly, for example, {voiced} stops are never [-voice], and {voiceless aspirated} stops are never [+voice]. And regardless of which phonetic pair is contrasted in a given case, the one implementing [+voice] is always phonetically more voiced than the one implementing [-voice]. For example, English speakers who do not prevoice often implement their phonological [±voice] contrast as {voiced} vs. {voiceless unaspirated} in some contexts, or as {voiceless unaspirated} vs. {aspirated}. Whichever pair is chosen, the more voiced one implements [+voice], and the less voiced one implements [-voice]. This system is illustrated in Figure 7 for nonalveolars in two representative contexts. In either case, the {voiceless unaspirated} category is used, sometimes as [+voice] and sometimes as [-voice].

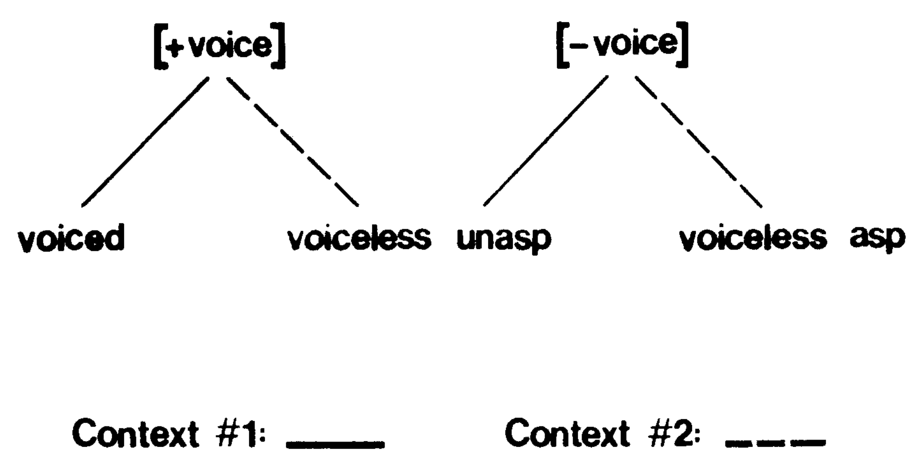


Figure 7. Schematic of phonetic categories used to implement the [voice] feature in different contexts.

Comparing English with Polish implementation, it appears that this diagram can also be taken to represent two different languages as well as two different environments. The phonological identity of the {voiceless unaspirated} category will be different in different languages, depending on the phonetic contrast used. It will be phonologically [+voice] if it is the left member of the [voice] contrast, and it will be phonologically [-voice] if it is the right member of the [voice] contrast. Thus the [voice] feature's + and - values can only be phonetically interpreted in a relative sense, as 'more' and 'less', not 'with' and 'without', voicing. Such a relative definition is also required by the somewhat abstract nature of the phonetic voicing dimension, whose physical characteristics will vary with position and context.

Depending on the phonetic implementation being used, the phonetic {voiceless unaspirated} category may be unused, may implement [+voice] stops, or may implement [-voice] stops. That is, it acts as a 'swing' category for phonological implementation. If it is known that a class of stops in a language have short lag VOT's, it cannot then be said whether they are [+voice] or [-voice]. However, this does not mean that there are no phonetic differences between [+voice] and [-voice] short lag stops, as we have already seen. It simply means that the possible phonetic differences are never employed contrastively.

3.3. Phonetic detail rules

The rules of phonetic category implementation are thus language-specific rules that draw from the universal set of phonetic categories. Consider now the rules that map the three phonetic categories into their concrete physical realizations. Are these rules language specific or universal? Superficial examination would suggest that the former must be the case. For example, compare the VOT distributions for short lag stops in Polish and English (Figures 1 and 3). It can be seen that the distributions for Polish are about 5 msec VOT higher than those for English, although both sets are clearly short lag VOT. We may have to state that Polish short lag stops are slightly more aspirated than English ones by means of slightly different quantitative rules for the two languages.

However, such a rule would not be necessary if the observed variation could be derived by a general principle. One such principle would be 'polarization' of two adjacent categories along the voicing dimension. According to this principle, within the limits of the implementation chosen, i.e. the phonetic categories, there is maximal separation of the distributions of values. In effect, this principle says that the 'dispersion' theory of Liljencrants and Lindblom (1972) operates not over a continuous phonetic space or dimension, but within the discrete categories of the phonology.

Under this account, the contrasts in Polish and in English are heightened through polarization, with a differential effect on voiceless unaspirated stops. Data on Spanish vs. English VOT can be interpreted in the same way. Ladefoged and Kim (ms) compared English /b/ with Spanish /b/ and /p/, looking for VOT differences. They found roughly that English /b/ covers a less extreme range of VOT values than the two Spanish categories combined -- that is, it includes lead values (but not as voiced) and lag values (but not as voiceless). They proposed a fourth phonetic category corresponding to the English /b/ distribution. However, these data can be handled with three categories and the polarization principle. Under this account, English /b/ has a bicategory distribution as lead and short lag VOT values, while Spanish /b/ and /p/ each are realized by a single VOT category. The English /b/ lag VOT values are then polarized away from the higher lag values implementing English /p/. The Spanish /b/ distribution is polarized away from Spanish /p/, towards long lead values, while the Spanish /p/ distribution is polarized in the opposite direction.

On the other hand, some of the data presented here and elsewhere on languages with initial contrasts of {voiceless unaspirated} vs. {voiceless aspirated} stops are less clearly explained by polarization. The English and German data presented above, and the similar data on Swedish in Keating, Linker, and Huffman (1983, this volume) show that even for similar data sets some differences in VOT are found for the same phonetic category implementations. What's more, English basically has more aspirated [-voice] stops even though, with greater likelihood of [+voice] {voiced} stops, there is less pressure for dispersion. The German medial pre-stress contrast is so marginal in terms of VOT that polarization does not seem to be at work there, either.

Data from more languages is clearly required to verify this hypothesis about distribution dispersion. While polarization may turn out to be a wrong hypothesis, the principle involved is an attractive one: surface differences across languages are accounted for, not by an elaborate set of phonological distinctions or language-specific phonetic rules, but by the interaction of phonetic mapping constraints and a universal phonetic principle. A similar case arises in the treatment of stress and positional allophones: to what extent do stress and position per se universally affect quantitative detail and even phonetic category implementation? For the beginnings of an answer, see Keating, Linker, and Huffman (1983, this volume).

The issue here is whether rules of category choice are the only phonetic rules for voicing in the grammar. If differences in the way a given category appears are found across languages with identical phonological systems, then the grammars of those languages will have to contain rather specific quantitative rules. Obviously, a fair degree of uniformity across similarly specified contrasts is predicted. However, it must be noted that such comparisons of distributions require data sets that are large enough to be reliable, and that have similar vowels following the consonants in question, since vowel identity can affect VOT. It is possible that at least some cross-language differences observable in the literature are due to failure to satisfy one or both of these requirements. In any event, universal principles must be considered before we resort to language-specific quantitative rules.

4. Neutralization and Markedness

The framework that has been proposed distinguishes a phonological from a phonetic level in order to characterize cross-language and cross-context differences and similarities adequately. Does this proliferation of formal levels, and the concomitant abstractness of the phonetic categories, play any other role in the grammar? I will now suggest that the formal, rather abstract phonetic level proposed here provides an additional advantage: that it is the correct level for statements about markedness phenomena, providing a new approach to an old problem.

Markedness phenomena have generally been treated as phonological, as a property of phonemes and/or phonological features (Trubetzkoy 1939, Jakobson 1962, Chomsky and Halle 1968, Kean 1975). So, for example, 'voiceless' stops are found to occur in more languages and in more environments than are 'voiced' stops, and voiceless stops are taken to be the unmarked category value. Languages without a voicing contrast will have voiceless, rather than voiced, stops. And, as Trubetzkoy pointed out, when neutralization is not contextual, voiceless, rather than voiced, stops are found in positions of neutralization (where neutralization means either phonological rules or defective distribution). The same sort of preference holds for unaspirated over aspirated stops, and those unaspirated stops are typically voiceless, but discussions of markedness rarely relate the voicing case to the aspiration case.

There is a simple physical generalization underlying the apparent phonological one: physically, more environments favor {voiceless unaspirated} stops over other categories. Westbury and Keating (1980, forthcoming) used a computer implementation of an electrical analog model of vocal tract aerodynamics to demonstrate that in absolute initial and final positions, the most natural vocal tract settings result in this category. (In intersonorant position they result in voiced stops, which are not unambiguously preferred over voiceless stops (Houlihan 1982)). We claimed that non-contextual neutralization reflects this phonetic preference, thereby resulting in the physically unmarked category. However, the final devoiced stops found in neutralization are not physically identical to the voiceless stops they are neutralized with. Westbury and Keating (1980) suggested that in Polish the underlyingly [+voice] stops have more closure voicing, and Dinnsen (1982) found that final neutralized stops in Catalan differ in closure duration. While these differences apparently are not audible, they do indicate that for speakers, underlying phonological feature values must be available to the physical detail rules. We will return to this point below. For now, note that since neutralization is not physically complete, it can only be captured by a somewhat abstract phonetic category system that equates as voiceless unaspirated two slightly different kinds of stops.

With this account, a reanalysis of a central part of Trubetzkoy's theory of markedness and neutralization presents itself. In Trubetzkoy's theory, if two phonemes differ in a single feature, the one which is articulatorily simpler is 'unmarked' and has the minus value of the feature in question (p. 146). (While the theory included other types of neutralization, this type is probably the best known.) However, this phonetic basis for markedness could be ignored by a particular language, in which case the plus value of the feature would appear to be unmarked, since the phoneme with that value would occur in positions of neutralization (p. 147). Nonetheless, Trubetzkoy tried to maintain the rather interesting correlation between the phonetically and phonologically unmarked feature values, going as far as using the correlation as a criterion for deciding some 'doubtful cases'. For example, 'in a language in which voiced lenis consonants form a neutralizable opposition with voiceless fortis consonants, and in which the archiphoneme in the positions of neutralization is represented by a voiceless fortis consonant, the correlation of voice is present' (p. 147), since voicelessness but not fortisness is phonetically unmarked. Similarly, if in a t - d opposition, the t is phonologically unmarked, then the relevant feature must be Voice, since t is voiceless, but if the d is phonologically unmarked, then the relevant feature must be Tension, since d is lax ([-tense]) (p. 76-7). That is, the tenseness feature bore the burden of maintaining the correlation between phonetic and phonological markedness in cases involving voicing. Otherwise, tenseness did little work: it was left phonetically vague (and has eluded phoneticians since), and it was largely redundant with voicing. Generalizations about voicing had to be repeated for tenseness (e.g. p. 80), much as we saw with Chen's vowel length data. Because neutralization was thought of as a phonological phenomenon, the situation with voicing encouraged positing this second feature to act as a mirror image of Voice, with a minus value (Lax) which would be phonetically voiced in certain cases.

The hypothesis that phonetically natural categories will often be unmarked and be more likely to occur is similar to the hypothesis formulated here. The difference is that the present proposal does not necessitate a pseudo-phonetic feature to represent the fact that phonological categories work differently in different languages. Instead, an additional level of representation permits markedness to be stated without regard for which phonological feature value the phonetically unmarked voiceless unaspirated category is implementing. A language will neutralize to that phonetic category regardless of which phonological value is entailed. Thus the multiplication of phonological features like Tension, to the extent that it is due to assumptions about markedness and neutralization, can be seen to be an unnecessary complication arising from Trubetzkoy's theory.

Furthermore, the separation of levels of representation allows statements about inventory preferences across languages to clearly distinguish phonological from phonetic preferences. Languages may in fact prefer certain phonological feature contrasts over others, but within those contrasts they appear to prefer particular phonetic categories as well. So, for example, while it may be said that languages will not have /g/ without /k/, it probably is the case that the preferred phonetic category is {voiceless unaspirated}, although for labials the preferred category may be {voiced} (Maddieson 1981).

5. Discussion

5.1. Comparison with other models

Let us return now to the question of how the proposals made here differ from, and improve upon, previous proposals. The present model is essentially an extension of the SPE model. Like SPE, it has a phonological component with feature-changing rules and feature-implementing rules, which provide non-binary values with some language-specific detail. However, the modified systematic phonetic level consists of phonetic categories along a phonetic dimension; the inventory of categories imposes constraints on possible contrasts that a simple scale does not. The phonetic categories necessitate less phonetic detail at both the phonetic and phonological levels of representation, but in this way allow generalizations at each level to be expressed.

Because this phonetic category representation is more abstract than SPE's systematic phonetic level, its phonetic transcription level is also more abstract than theirs. Various subtle phonetic differences that would distinguish one dialect or language from another are not represented. The phonetic category level is designed to be the one level that phoneticians, including Halle and Stevens, refer to constantly: the level of possible contrasts. Until now, phoneticians have had to surreptitiously use the surface phonemic level when they talk about constructs such as dispersion of categories in phonetic space, or freedom of allophonic variation. With the phonetic category level, one can refer directly to those categories that a language uses in its phonetic realizations.

The alternative to such an intermediate level would appear to be a system like Pierrehumbert (1980) proposes for intonation, in which phonological values are mapped directly into continuous (quantitative) values. While this proposal is intriguing, Pierrehumbert herself suggests that segmental features would have an additional representation corresponding to a phonetic transcription. In the case of [voice], mapping directly into quantitative values would fail to express the severe limitations on possible phonetic contrasts. Presumably, an attempt to make those constraints explicit would amount to a system of the sort proposed here.

Another aspect of Pierrehumbert's proposals, however, is supported under our system. Pierrehumbert argues that the phonetic rules that derive pitch contours must not replace the phonological values (H or L) of the tonal system, but simply further specify them with phonetic values. This is because certain rules of Pierrehumbert's system must refer simultaneously to a tone's phonetic value and to its phonological category. A similar case is found in the phonetics of voicing in Polish. Such a corepresentation of phonological and phonetic values is required for the proper treatment of two phenomena already mentioned, final (non-neutralizing) devoicing and fundamental frequency control. In both cases we saw slight physical differences between members of the 'same' phonetic category, {voiceless unaspirated}. That such differences should not be expressed by having two categories follows from the fact that no language uses these kinds of voiceless unaspirated stops contrastively: where Polish does use two, in final position, no contrast is perceived by native speakers. The differences correlate with the phonological feature value for [voice]. Therefore phonetic implementation cannot replace phonological values with phonetic category values. Rather, it must add phonetic values without removing phonological values, so that both kinds will be available at the point of physical (quantitative) implementation. At that point the co-occurrence of, for example, [+voice] and {voiceless unaspirated} will result in a lowered fundamental frequency at voice onset, or a larger amount of closure voicing in a final stop. In this sense, then, neutralizing rules will not be formally identical to the usual kind of feature changing rule (Port, Mitleb, and O'Dell 1982). The phonological value of [voice] must also be available for the duration specification of a preceding vowel, as we have seen.

Consider briefly what sort of grammatical organization is implied by these proposals. Underlying phonological feature values ([±voice]) may be transformed by phonological rules into opposite values. Then the language-specific phonetic implementation rules select major phonetic categories corresponding to the phonological values, according to context (including adjacent segments) and possibly following general trends. The two representations, however, are present simultaneously. In the case of [voice], it has been convenient to think of the VOT dimension as being represented in the implementation rules. However, more properly these rules will convert representations as [±voice] to the three basic phonetic voicing categories, for which VOT values are an approximation. These phonetic categories must then be converted to values along physical scales, in ways which are sensitive both to phonological value and to context, place of articulation, etc. The scales can be thought of as acoustic dimensions such as VOT and closure duration (with more than one scale having to be specified for a given set of categories), or they can be thought of as articulatory specifications controlling actual speech production mechanisms. Whether this conversion can be accomplished entirely by universal rules of phonetics, or requires additional language-specific rules, remains an open question, although limited evidence suggests that at least some aspects of conversion can be universally predicted.

5.2. Summary and Conclusions

To summarize, I have tried to show that certain cross-language phonetic differences can best be expressed as differences in the realization or implementation of phonological feature contrasts as phonetic categories. A phonological feature, [±voice], was proposed to account for rule equivalence across languages with phonetically different contrasts, justifying the use of the same symbols for different sounds. This system is motivated by the allophonic variation found across contexts within a single language such as English. Three phonetic categories, {voiced}, {voiceless unaspirated}, and {voiceless aspirated}, were proposed to express the maximum number of contrasts found along the voicing dimension, and to describe the markedness relation among those categories. Evidence was presented that the boundaries defining the three categories derive from physiological, largely auditory, constraints. It was also proposed that subtle differences between languages in exactly where between the category boundaries the categories lie could be accounted for by principles such as polarization, which separates categories within the limits imposed by the boundaries.

The analysis presented is in accord with a view that languages do not differ without limit; rather, variation is constrained at different levels in the grammar. In the course of this account, several empirical claims were made that must be tested further. One is that languages with different phonetic implementations will nonetheless share the same phonological rules. Another is that contrastive voicing categories are restricted to certain parts of the phonetic voicing dimension, respecting the category boundaries imposed by universal phonetics. The polarization hypothesis must also be tested. Still another claim is that certain phonetic categories may be in a sense 'ambiguous', receiving minutely different physical realizations and inducing different contextual effects on the basis of their phonological category. More generally, it is posited that a level of representation between that of phonological features and the output of phonetic detail rules, a level that to some extent abstracts away from phonetic detail, is useful and necessary.

We have seen that while universal phonetics provides a constrained set of phonetic voicing categories, various languages use all possible combinations in their implementations of [±voice]. The choice of implementation rules must be specified for each context in each language, since there does not seem to be any way of predicting categories across environments, even though there are only three main possibilities to choose from. Further, we have seen that the need for language-specific quantitative rules is an open question, depending on the viability of hypotheses such as the polarization principle. The answer to this question must await more data on slight differences between languages with the same implementation of a contrast, so that principles like polarization can be proposed and tested. To the extent that such principles are insufficient to account for observed variation, even low-level phonetic rules will be seen to be language-specific. In that case, the role of universal phonetics would largely be to constrain the form and substance of such rules, in part by establishing the phonetic category boundaries that would limit possible cross-language differences. Such constraints would obviously contribute to the learnability of the phonological and phonetic systems.

Rules of implementation and of detail must eventually be provided for [voice] in other kinds of segments, as well as for other phonological features, if this model is to be supported further. The success of this enterprise will depend on finding evidence for a limited number of major phonetic categories available for each feature. The extent of psychoacoustic investigations of phonetic dimensions suggests that we will not progress far in the near future. Nonetheless, hypotheses such as those of Stevens and Blumstein (1981) on the acoustic properties underlying various phonological features, coupled with studies of the distribution of major categories, could prove rewarding. Promising starting points for further research include dental vs. alveolar places of articulation (Lahiri and Blumstein 1981), and different kinds of flaps (Price 1981).

It is assumed by phonologists that the phonological component of the grammar is highly structured, and that its formal properties as well as its substance are of linguistic interest. It might be thought that, in contrast, the organization of phonetics is relatively trivial, comprising a physical scale for each phonological feature and, for each language, a value along the scale at which it is divided into phonological categories. In this paper I have tried to show that such a simple system will not work in the case of [voice] contrasts. The phonetic rules developed to account for the data presented, and the phonetic level of representation derived by them, have interesting properties, suggesting that the phonetic component of the grammar is more structured and richer than has been supposed. It is to be hoped that further work along these lines will clarify the place of phonetics in a grammar of linguistic competence.

ACKNOWLEDGMENTS

This research was supported in part by a postdoctoral fellowship at MIT from NIH, by a research grant from UCLA, and by an NSF grant to Peter Ladefoged. I would like to thank Elan Dresher, Louis Goldstein, Bruce Hayes, Peter Ladefoged, Ian Maddieson, John Ohala, Patti Jo Price, Stephanie Shattuck-Hufnagel, and John Westbury for comments and/or advice, Don Brusasco, Marie Huffman, Ellen Jackson, and Wendy Linker for data collection, and Phil Lieberman and Sheila Blumstein for the initial idea. While not all of them agree with me, they have all helped in some way.

FOOTNOTES

[1] Chen's study shows at least this, although two caveats are in order. First, not all languages do show this effect (Flege (1979), Keating (1979)). Second, Chen's claim that different languages show different degrees of lengthening has some problems. It is known that there is less lengthening in disyllables than in monosyllables, yet Chen's samples were not matched for number of syllables. Some of Chen's "language" effects are surely sample effects. However, Mack (1982) compared just French and English with carefully matched word lists, and did find a difference of this sort, so Chen's claim may still prove true.

[2] A lag VOT value reflects how long it takes the vocal cords to start vibrating; a long lag VOT value indicates that the cords were quite separated at release and took some time to get back together. Aspiration is friction noise generated at the still-open glottis by the flow of air through the vocal tract after stop release. Therefore as long as the vocal cords are apart (the longer the lag), air can flow between them. It is in this sense that VOT is a measure of aspiration.

[3]Other unpublished data of Leigh Lisker and Patti Price suggests a similar situation for English. Since {voiceless unaspirated} stops typically have one or two pitch periods of voicing at the beginning of closure, and {voiced} stops are often not voiced throughout the closure and release, it is not obvious what it means to be listening for 'closure voicing', however.

[4]In Lisker and Abramson (1967), speakers were shown to produce more prevoicing in minimal pair readings, which were assumed to be a more careful style. However, the opposite was found by Flege and Massey (1980), who also found that speakers prevoiced the same amount as a session progressed. (They had thought that as speakers relaxed their more casual speech would have less prevoicing.) In support of this latter interpretation of the relation of prevoicing to carefulness of speech, Malsheen (1980) showed that mothers do not prevoice more in speaking to their children, although they do aspirate more strongly, and Chen et al. (1980) showed that speakers do not prevoice more in speaking 'clearly' to deaf listeners. Other factors influencing stop prevoicing include the phonetic context (place of articulation, following vowel) and the speaker's sex (Smith and Westbury 1975).

[5]The most interesting part of their study was their use of three VOT continua, one at each of the places of articulation labial, alveolar, and velar. This allowed them to test the known effect with adults for place of articulation on VOT boundaries. The chinchillas' boundaries did in fact vary according to place of articulation as the humans' did, from about +25 msec for labials to about +42 msec for velars. However, their identification functions were less steep. Such sharpening of category boundaries with linguistic experience is not unique to the human/animal comparison. An experiment in which both Czech and American listeners divided a continuum of Czech words varying in their vowel duration between a phonemic short vowel and long vowel pair showed the same kind of result. The cross-over points for the Czechs (who were performing a linguistic task) and the Americans (who were performing a non-linguistic task) were the same, but the boundaries were much steeper for the Czechs (Keating 1978).

[6]Such as place of articulation, vowel context, and speaker's sex.

[7]The result that VOT value for initial [-voice] stops does not depend on the stress of the following vowel would seem to refute the claim in the literature (e.g. Hoard 1971, Kahn 1976) that aspiration correlates with stress. However, that claim might still be correct. The high VOT values obtained for initial [-voice] stops in stressless syllables are due to reduction and devoicing of the entire stressless syllable. That is, the high VOT values are not, strictly speaking, the same as aspiration of the initial consonant. This reduced syllable devoicing does not necessarily require a special mechanism. Lack of stress could cause shortening of the syllable, while a minimum time interval could be required to bring the vocal cords into position for vibration; the syllable would fall entirely within this interval and so would be voiceless.

[8]William Moulton informs me that German speakers are taught to always aspirate [-voice] stops in Standard German, and that therefore the failure of this speaker to deaspirate, and even to voice the [+voice] stops, medially, may be due to an artificial style of speech.

REFERENCES

Abraham, Arthur S. 1972. Word-final stops in Thai. Tai phonetics and phonology, ed. by Jimmy G. Harris and Richard B. Noss, 1-7. Bangkok: Central Institute of English Language.

Abramson, Arthur S., and Leigh Lisker. 1972. Voice-timing perception in Spanish word-initial stops. J. Phon. 1.1-8.

Anderson, Stephen R. 1978. Tone Features. Tone: A Linguistic Survey, ed. by Victoria A. Fromkin, 133-176. New York: Academic Press.

Carney, Arlene Earley, Gregory P. Widin, and Neal F. Viemeister. 1977. Noncategorical perception of stop consonants differing in VOT. J. Acoust. Soc. Am 62.961-970.

Chen, Matthew. 1970. Vowel length variation as a function of the voicing of the consonant environment. Phonetica 22.129-59.

Chen, F., V. Zue, M. Picheny, N. Durlach, and L. Braida. 1980. Speaking clearly: Acoustic characteristics and intelligibility of stop consonants. Paper presented at 99th meeting of Acoust. Soc. Am.

Chomsky, Noam, and Morris Halle. 1968. The Sound Pattern of English. New York: Harper and Row.

Dent, Laurel. 1976. Voice onset time of spontaneously spoken Spanish voiceless stops. Paper presented at 91st meeting of Acoust. Soc. Am.

Dinnsen, Daniel. 1982. Abstract Phonetic Implementation Rules and Word-Final Devoicing in Catalan. Paper presented at Annual Meeting of LSA, San Diego.

Donald, Susan Lea. 1978. The Perception of Voicing Contrasts in Thai and English. Unpublished dissertation, University of Connecticut at Storrs.

Dow, Francis. 1972. An Outline of Mandarin Phonetics. Canberra: Faculty of Asian Studies in Association with the Australian National University Press.

Flege, James E. 1979. Phonetic Interference in Second Language Acquisition. Unpublished dissertation, Indiana University.

Flege, James E. 1982. Laryngeal timing and phonation onset in utterance-initial English stops. J. Phon. 10.177-92.

Flege, James, and William Brown. 1982. Effects of utterance position on English speech timing. Phonetica. (In press)

Flege, James, and Karen Massey. 1980. English prevoicing: Random or controlled?. Paper presented at the Summer meeting of the Linguistic Society of America.

Fischer-Jørgensen, Eli. 1954. Acoustic Analysis of Stop Consonants. Miscellanea Phonetica, Vol. II. 42-59.

Fox, Robert A., and Dale Terbeek. 1977. Dental flaps, vowel duration and rule ordering in American English. J. Phon. 5.27-34.

Goldsmith, John. 1976. Autosegmental Phonology. MIT Dissertation. Distributed by Indiana University Linguistics Club.

Greenberg, J. H. 1970. Some generalizations concerning glottalic consonants, especially implosives. IJAL 36.123-145.

Haas, Mary R. 1956. The Thai System of Writing. Washington, D. C.: American Council of Learned Societies.

Halle, Morris, and Kenneth N. Stevens. 1971. A note on laryngeal features. Quarterly Progress Report of the Research Laboratory of Electronics, M.I.T. 101.198-213.

Hayes, Bruce. 1983. The phonetics and phonology of Russian voicing assimilation. Unpublished ms, UCLA.

Hebert, R. and N. Poppe. 1963. Kirghiz Manual. Indiana University Publications, Uralic and Altaic Series, 33. Indiana University, Bloomington.

Heffner, R-M. S. 1969. General Phonetics. Madison: The University of Wisconsin Press.

Hirsh, Ira. 1959. Auditory Perception of Temporal Order. J. Acoust. Soc. Am. 31.759-767.

Hirsh, Ira. 1975. Temporal Aspects of Hearing. The Nervous System, Vol. 3: Human Communication and Its Disorders, ed. by D. B. Tower, 157-162. New York: Raven Press.

Hoard, James. 1971. Aspiration, tenseness, and syllabication in English. Lg. 47.133-40.

Hombert, Jean-Marie, John J. Ohala, and William G. Ewan. 1979. Phonetic Explanations for the Development of Tones. Lg. 55.37-58.

Houlihan, Kathleen. 1982. Is Intervocalic Voicing a Natural Rule? Paper presented at Annual Meeting of LSA, San Diego.

Hyman, Larry M. 1975. Phonology: Theory and Analysis. New York: Holt, Rinehart, & Winston.

Jakobson, Roman. 1962. Selected Writings, Vol I: Phonology. The Hague: Mouton.

Jakobson, Roman, Gunnar Fant, and Morris Halle. 1952. Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates. Cambridge: MIT Press.

Kahn, Daniel. 1976. Syllable-Based Generalizations in English Phonology. MIT Dissertation. Distributed by Indiana University Linguistics Club.

Kean, Mary-Louise. 1975. The Theory of Markedness in Generative Grammar. MIT Dissertation. Distributed by Indiana University Linguistics Club.

Keating, Patricia A. 1978. Perception of vowel length in Czech and English. Paper presented at the annual meeting of the LSA.

Keating, Patricia A. 1979. A phonetic study of a voicing contrast in Polish. Unpublished dissertation, Brown University.

Keating, Patricia A., Wendy Linker, and Marie Huffman. 1983. Patterns in Allophone Distribution for Voiced and Voiceless Stops. This volume.

Keating, Patricia A., Michael J. Mikoš, and William F. Ganong III. 1981. A cross-language study of range of voice onset time in the perception of initial stop voicing. J. Acoust. Soc. Am. 70.1261-71.

Kuhl, Patricia K. 1978. Predispositions for the Perception of Speech-Sound Categories: A Species-Specific Phenomenon?. Communicative and Cognitive Abilities -- Early Behavioral Assessment, ed. by Fred D. Minifie and Lyle L. Lloyd, 229-256. Baltimore: University Park Press.

Kuhl, Patricia K., and James D. Miller. 1975. Speech perception by the chinchilla: Voiced - voiceless distinction in alveolar plosive consonants. Science 190.69-72.

Kuhl, Patricia K., and James D. Miller. 1978. Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. J. Acoust. Soc. Am. 63.905-17.

Ladefoged, Peter. 1978. Phonetic differences within and between languages. UCLA WPP 41.32-40.

Ladefoged, Peter. 1982. A Course in Phonetics. 2nd edition. New York: Harcourt Brace Jovanovich.

Ladefoged, Peter, and Chin W. Kim. 1965. Comments on 'Voicing in Initial Stops'. Unpublished ms, UCLA.

Lahiri, Aditi, and Sheila E. Blumstein. 1981. A reconsideration of acoustic invariance for place of articulation in stop consonants: Evidence from cross-language studies. Paper presented at the 102th meeting of Acoust. Soc. Am.

Lieberman, Philip. 1970. Towards a unified phonetic theory. Linguistic Inquiry 1.307-22.

Lieberman, Philip. 1977. Speech Physiology and Acoustic Phonetics. New York: Macmillan Publishing.

Liljencrants, J., and Bjorn Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. Lg. 48.839-862.

Lisker, Leigh, and Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. Word 20.384-422.

Lisker, Leigh, and Arthur S. Abramson. 1967. Some effects of context on voice onset time in English stops. Lang. Speech 10.1-28.

Lisker, Leigh, and Arthur S. Abramson. 1971. Distinctive features and laryngeal control. Lg. 47.767-85.

Lisker, Leigh, Arthur S. Abramson, Franklin S. Cooper, and Malcolm H. Schvey. 1969. Transillumination of the Larynx in Running Speech. J. Acoust. Soc. Am. 45.1544-46.

Lofqvist, Anders. 1980. Interarticulator programming in stop production. J. Phon. 8.475-490.

Mack, Molly. 1982. Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. J. Acoust. Soc. Am. 71.173-78.

Maddieson, Ian. 1977. Further studies on vowel length before aspirated consonants. UCLA WPP 38.82-90.

Maddieson, Ian. 1981. UPSID: Data and Index. UCLA Working Papers in Phonetics 53.

Malsheen, Bathseba J. 1980. Two Hypotheses for Phonetic Clarification in the Speech of Mothers to Children. Child Phonology, Vol. 2: Perception, ed. by Grace H. Yeni-Komshian, James F. Kavanagh, and Charles A. Ferguson. New York: Academic Press, 173-184.

Mikoś, Michael J. 1977. Problems in Polish Phonology. Unpublished dissertation, Brown University.

Miller, J. D., C. C. Wier, R. Pastore, W. J. Kelly, and R. J. Dooling. 1976. Discrimination and labeling of noise - buzz sequences with varying noise-lead times: An example of categorical perception. J. Acoust. Soc. Am. 60.410-17.

Pierrehumbert, Janet Breckenridge. 1980. The Phonology and Phonetics of English Intonation. Unpublished dissertation, M.I.T.

Pisoni, David B. 1977. Identification and discrimination of the relative onset of two component tones: Implications for the perception of voicing in stops. J. Acoust. Soc. Am. 61.1352-61.

Port, Robert, F. Mitleb, and M. O'Dell. 1982. Neutralization of obstruent voicing in German is incomplete. Paper presented at the 102nd meeting of the Acoustical Society of America.

Price, Patti Jo. 1981. A cross-linguistic study of flaps in Japanese and in American English. Unpublished dissertation, U. Penn.

Smith, Bruce, and John R. Westbury. 1975. Temporal control of voicing during occlusion. Paper presented at the 89th meeting of the Acoustical Society of America.

Stevens, Kenneth N., and Sheila E. Blumstein. 1981. The Search for Invariant Acoustic Correlates of Phonetic Features. Perspectives on the Study of Speech, ed. by Peter D. Eimas and Joanne L. Miller, 1-38. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Stevens, Kenneth N., and Dennis H. Klatt. 1974. Role of formant transitions in the voiced - voiceless distinction for stops. J. Acoust. Soc. Am. 55.653-59.

Streeter, Lynn A. 1976. Kikuyu labial and apical stop discrimination. J. Phon. 4.43-9.

Trubetzkoy, Nikolai S. 1939. Principles of Phonology. Translated by Christiane A. M. Baltaxe, 1969. Berkeley and Los Angeles: University of California Press.

Vanvik, A. 1972. A Phonetic-Phonemic Analysis of Standard Eastern Norwegian, Part I. Norwegian Journal of Linguistics 26.1-64.

Waters, R., and W. Wilson.1976. Speech perception by rhesus monkeys: the voicing distinction in synthesized labial and velar stop consonants. Perc. Psychophys. 19.285-89.

Weismer, Gary. 1980. Control of the voicing distinction for intervocalic stops and fricatives: some data and theoretical considerations. J. Phon. 8.427-38.

Westbury, John R. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. J. Acoust. Soc. Am. In press.

Westbury, John R., and Patricia A. Keating. 1980. A model of stop consonant voicing and a theory of markedness. Paper presented at the Linguistic Society of America.

Westbury, John R. and Patricia A. Keating. Forthcoming. Unpublished manuscript, revised and expanded version of Westbury and Keating (1980).

Williams, Lee. 1974. Speech Perception and Production as a Function of Exposure to a Second Language. Unpublished dissertation, Harvard University.

Yoshioka, Hirohide, Anders Lofqvist, and Hajime Hirose. 1981. Laryngeal adjustments in the production of consonant clusters and geminates in American English. J. Acoust. Soc. Am. 70.1615-23.

Patterns in allophone distribution for voiced and voiceless stops

Patricia Keating, Wendy Linker and Marie Huffman

ABSTRACT

Do languages prefer certain types of stop consonants in certain environments? A survey of 51 languages reveals two general trends. First, many languages limit the occurrence of final stops; final devoicing is one instance of this phenomenon. Second, other variation is largely limited to languages whose contrasts in initial position involve short lag vs. long lag VOT, that is, aspiration contrasts. Among these languages, many sorts of variation are found, which therefore must be described by language-specific rules.

It has often been observed that languages prefer voiceless over voiced consonants. At the same time, however, it is generally thought that intervocalic consonants will most naturally be voiced. If this is indeed the case, then we should see either consistent patterns of contrast neutralization, or of allophonic variation, with voiceless consonants being found everywhere except intervocalically. Work with an aerodynamic model (Westbury and Keating 1980, forthcoming) leads one to expect just that: voiceless unaspirated initial and final stops but voiced medial stops. Nonetheless, data from natural languages does not appear to lend overwhelming support to these predictions, particularly in terms of medial voicing (Houlihan 1982). It became clear to us that better data on positional variation in consonant voicing was required to compare with results from physiological modeling studies. Therefore we began a survey of the allophones of voiced and voiceless stop phonemes in various languages.

Allophone survey

The languages surveyed include those in Lisker and Abramson (1964)'s study of word-initial VOT values, and the various languages whose initial VOT values have been similarly determined since then. These languages form less than half the database. Additional languages were added to increase the overall variety, both genetic and phonetic, of the database, and to provide sufficient examples of phenomena of particular interest. All stop consonants, including glottalic segments, were included in the survey. The data consist primarily of statements about allophone and phoneme distribution, largely derived from impressionistic transcriptions used for classical phonemicizations. In many cases the sources consulted are those used in compiling the UCLA Phonological Segment Inventory Database (UPSID, Maddieson 1981), a listing of the phonetic feature description of one principle allophone for each of the surface phonemes in 317 languages. UPSID was designed precisely to abstract away from the kind of positional variation that we are interested in, but its sources were often well suited to answering our questions. However, just as often such sources do not consider a sufficiently wide range of environments, and so where possible we supplemented information from the literature with our own or others' acoustic measurements, and with information from UCLA student phonetic projects and field notes. The major sources consulted and the allophones that were found are described for each language in the Appendix. Here we will describe the relevant trends observed across these data.

A major, though expected, trend across languages is the overwhelming preference for voiceless unaspirated stops. All of the languages surveyed except Bobo, Breton, and possibly Yidiɲ use at least this category.[1] Languages with a [voice] contrast generally use the voiceless unaspirated category in opposition to one or more other categories; exceptions include various languages with a [b] but no [p]. Beyond this preference, variation across positions in a word is also found. In the discussion that follows, we will focus on two kinds of variation: first, restrictions on final consonants, which are found in languages with all kinds of voicing contrasts, and restrictions on medial consonants, which hold largely of languages with initial aspiration contrasts.

Languages limit the consonants that can occur in final position to a greater or lesser extent. Some languages simply disallow final consonants, at least phonetically: Alyawarra (with a consistent final [ə]), Bobo, Kaititj, Kikuyu, Tiwi, marginally Hawaiian, with a final [ʔ], and Swahili, with final consonants only in loans. Other languages restrict final consonants to sonorants of one sort or another: Akan, Japanese, Mandarin, Tamil, Tswana, and Yidiɲ. Hausa allows final sonorants, and rarely [s], as well as having phonetic [ʔ] after short vowels. Spanish limits its final consonants to continuants. More common, however, is that stops are allowed but restricted in their manner. So, for example, Tagalog typically has unreleased final stops; English and Choctaw have optionally unreleased final stops. Burmese allows final nasals and a phonemic glottal stop /ʔ/ which in phrases may have various assimilatory phonetic realizations. Probably the best known type of final stop restriction is that final stops must be voiceless: Basque, Bulgarian, Cantonese, Choctaw, Dutch, Efik, Ewondo, Finnish, Gaelic, German, Polish, Russian, and Zoque. The voiceless stop restriction may, but need not, involve a synchronic phonological rule of final devoicing. It should be noted, however, that the phonological domain varies: in German it is syllable-final stops that are voiceless, while in Polish it is phrase-final stops. (Word-final but phrase-medial Polish voiceless stops can actually voice in some dialects (Mikoś 1977).) A few languages have only unreleased <u>and</u> voiceless final stops: Korean, Nama, Thai, Tikar, Vietnamese. Thus it can be seen that 'final devoicing' is just part of a larger trend to fewer segments and contrasts in final position. Other languages surveyed either have the same segments finally as initially and medially, or variation that involves, e.g. spirantization as part of the phonological contrast. Note that languages with four contrasting categories, e.g. Bengali, Hindi, Marathi, generally allow them all in final position, although not always in colloquial speech; Armenian and Tzeltal with three categories also allow them all in final position. Thus, while there is clearly a trend towards fewer contrasts in final position (involving other features such as place of articulation as well), and some variation in the allophones of the contrasting phonemes, it is certainly possible to sustain as many contrasts finally as elsewhere.

The situation for medial position is quite different, with variation being largely limited to one particular kind. Recall that one motivation for this study was an interest in the occurrence of medial voiced stops. If phonetically voiced stops are found in initial position, it is almost always in contrast with some other category. In this case, a 'preference' for voiced stops in medial position would have to involve neutralization of the contrast to a voiced variant. As noted before by Houlihan and Iverson (1979), languages typically do not have neutralization in medial position. On the other hand, if voiced stops do not occur in initial position, they could be found in medial position through allophonic variation in some category that does occur initially. First, a language with no voicing contrast could have voiced stops medially but voiceless stops elsewhere. We would in fact expect this to be the general pattern for languages without a contrast. Second, a language with an initial contrast between phonetically voiced and voiceless unaspirated stops could have a rule that changed the medial [+voice] stops into, e.g., fricatives, and [-voice] stops into voiced stops. Third, a language with an initial contrast between voiceless unaspirated and aspirated stops (e.g. most Germanic and Chinese languages) could have a rule that voiced one of those categories in medial position. Let us examine the data for each of these possible sources of medial voiced stops.

A reasonable hypothesis would be that languages without a [voice] contrast should show more allophonic variation than languages with a contrast, since a contrast might be thought to constrain possible variation within the phonetic space. Of the six languages without a contrast in our sample, only one is described as having a phonetic difference between initial and medial position: Yidiɲ appears to have systematic medial voicing, as predicted. However, none of the other five are described in this way. Spectrograms from one speaker of each language indicate that in Hawaiian medial stops are less aspirated than initial stops, but that in Alyawarra (an Arandic language of Australia) medial stops are if anything slightly more aspirated than initial stops. Kaititj (related to Alyawarra) shows no systematic differences, and Tiwi (also in Australia) is explicitly described as having no intervocalic voicing. Nama (a Khoisan language of Africa) voices /t/ and /k/ after rising tones, but these segments only occur word-initially; labials, which do occur medially and finally as well, do not undergo this change. Overall, then, these languages show surprisingly little allophonic variation, and what there is, is not consistent across the six languages.

To this number can be added four languages with rather marginal voicing contrasts. Finnish's only voiced stop, /d/, occurs only medially, and even then only as a product of gradation rules. In addition, Suomi (1980) describes Finnish /p t k/ as often 'voiced' intervocalically, but his own measurements show only a small amount of closure voicing. Next, the contrast in Kikuyu involves prenasalization as well as voicing, and there is no difference described between initial and medial positions. Third, in Cuna the initial voiceless unaspirated [p t k] are not part of a [voice] contrast, but in medial position they become voiced and contrast with voiceless geminates [p: t: k:]. Similarly, Tamil has no initial contrast but a medial contrast of short voiced stops (or fricatives) and geminate voiceless stops. Lastly, Choctaw's [voice] contrast is found only for labials; for all of its places of articulation, medial stops before unstressed vowels can be voiced. Again, then, there is some evidence for a preference for medial voiced stops, but no universal trend.

Languages with three or more contrasting categories in initial position often have the same categories in medial position (Bengali, Hausa, Marathi, Thai). However, Korean is a classic example of allophonic medial voicing, with initial [p t k] but medial [b d g]. E. Armenian is described as having a slight phonetic difference, with /b d g/ being more voiced medially, and possibly finally, than initially. Tikar has allophonic prenasalization, and Tzeltal has allophonic spirantization, of its medial [+voice] stops, with the other categories unchanged. Interestingly, Swahili is described as being in the process of eliminating its aspiration contrast by associating aspiration with initial position and with stress, a correlation that will re-appear below. Finally, word-initial voiceless unaspirated and aspirated stops in Burmese can become voiced in phrases, especially after a 'weakened' (toneless and reduced) syllable. The initial consonant in the weakened syllable is then itself eligible for voicing. That is, Burmese voicing is a phrasal rule, somewhat like Polish voicing of word-final stops before a vowel-initial word.

Languages with two contrasting categories in initial position, one voiced and one voiceless, and neither of which is aspirated, are languages with traditional 'voicing' contrasts (Arabic, Basque, Bulgarian, Dutch, Efik, Ewondo, French, Japanese, Polish, Russian, Spanish, Tagalog, Zoque). Their VOT values in initial position, where ascertained, are seen to be lead vs. short lag (Yeni-Komshian et al. 1977, Lisker and Abramson 1964, Caramazza and Yeni-Komshian 1974, Shimizu 1979, Keating et al. 1981, unpublished UCLA data). Of these languages, only Efik, Ewondo and Spanish show any real differences between initial and medial positions, and these differences do not just involve voicing. Efik medial alveolars become flaps, and Ewondo [+voice] stops may become continuants, as they do in Spanish. However, even in these cases where the [+voice] stops become something else, the [-voice] stops remain voiceless and unaspirated (cf. Dent 1976 on Spanish). Overall, there is little variation across these positions in these languages, and no tendency towards medial voicing, devoicing, or aspiration.

In contrast, languages whose initial stop contrast involves aspiration have been found to contrast short lag with long lag VOT values (Lisker and Abramson 1964, Fischer-Jørgensen 1954, Keating 1983). These languages generally do show differences between initial and medial position. The most common pattern across languages involves medial deaspiration of initial voiceless aspirated stops, and/or medial voicing of initial voiceless unaspirated stops. Degree of stress is often implicated in descriptions of this pattern: English, Gaelic, Hindi, Kirghiz, Mandarin, and Swedish, and sometimes German, are described as having unaspirated [-voice] stops before unstressed vowels. Mandarin, and sometimes German, are described as having voiced [+voice] stops before unstressed vowels. Choctaw, English, Kirghiz, Norwegian, Persian, Swedish, and Yidiɲ are all described as having voiced [+voice] medial stops, but not specifically before unstressed vowels. Danish treats medial stops differently, with spirantization of the [+voice] stops and voicing of the [-voice] stops. Akan and Cantonese are exceptions in that they appear to have no variation across positions.

In many of these cases, a sufficiently wide range of environments was not covered in the sources for us to be sure of the importance of stress as a conditioning factor. To clarify just two such cases, we collected acoustic data on English and Swedish, both languages with initial aspiration contrasts. While these are both Germanic languages, we emphasize that the phenomena of interest are not limited to such languages, as can be seen from the examples just given.

Acoustic data

Because a correlation between stress and phonetic detail was strongly indicated, we systematically varied both position in the word, that is, initial vs. medial, and stress, that is, main stress vs. other degrees of stress on a following vowel, for stops in each language. Some half dozen real words were used for each position + stress condition. All stops occurred before low vowels, but the vowel before each stop was not controlled. Words were read in isolation, so all initial stops were post-pausal. Six speakers of each language read their list once in a sound-treated booth. The items were not randomized, so as to make implicitly clear the desired stress contour. However, neither the experimental goals nor the lists were described to or discussed with the speakers.

Three measurements were made for each stop, shown in Figure 1. The top picture is a real waveform of the sort measured in a computer display; it includes some typical low-frequency noise that must be ignored. Below the waveform is a schematic version of the type we will use in presenting our results, and on the bottom is an indication of the measurements we made. The first was the duration of any voicing during closure, the second was the duration of voicelessness during closure, where that could be measured from an acoustic display. The sum of these two measures is the total closure duration. The third measure was the duration of voicelessness after the release, that is, lag VOT.

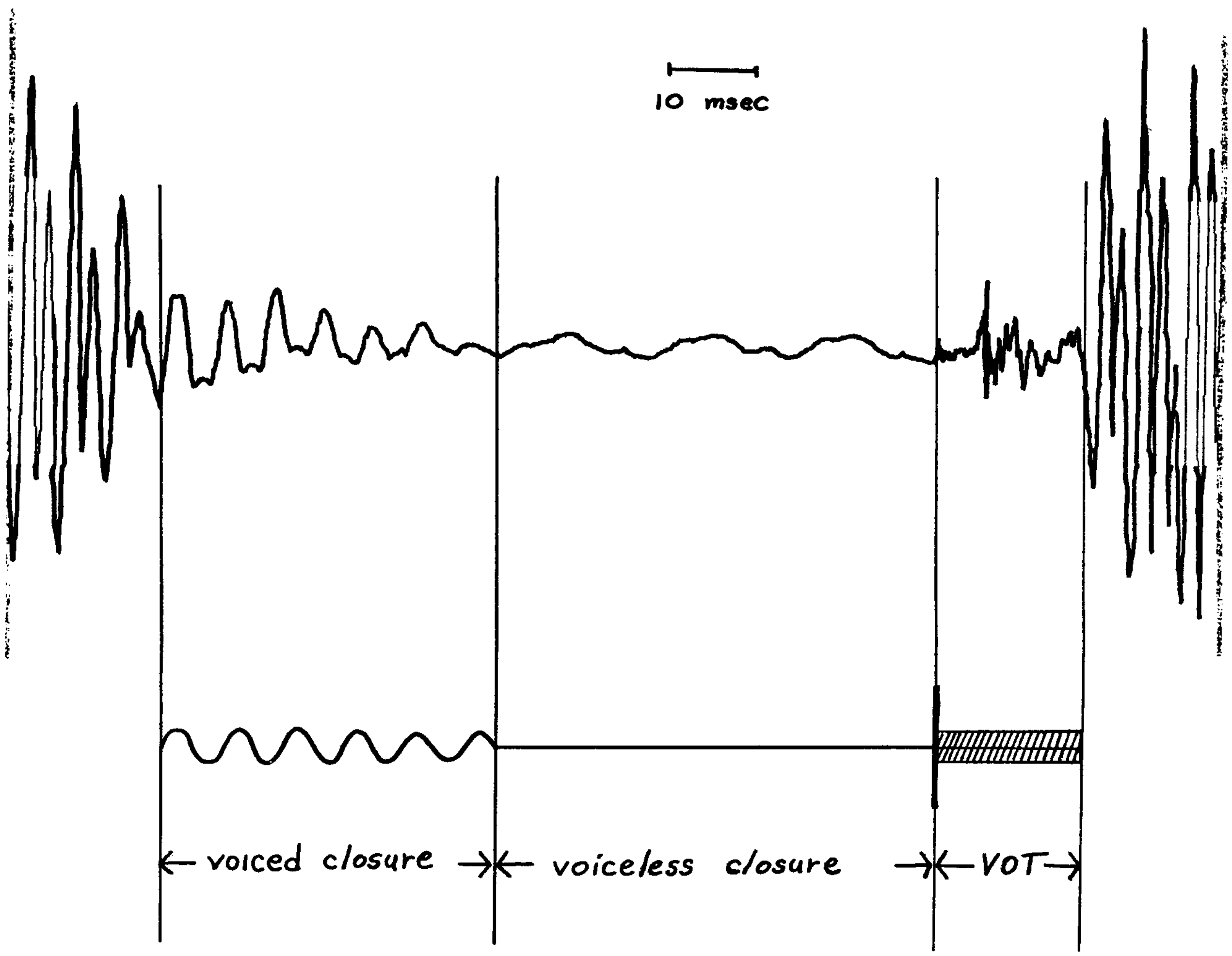


Figure 1. Acoustic measurements made for English and Swedish stops.

Results for English are shown in Figure 2. This experiment is similar to others in the literature (e.g. Lisker and Abramson 1964, 1967; Flege and Brown 1982), but it has the advantage of using the same speakers in a variety of contexts. For our speakers, while initial /b d g/ are generally voiceless and unaspirated, some speakers sometimes prevoice. Second, medial /p t k/ before reduced vowels generally have much lower VOT's, that is, are unaspirated. At the same time, medial /b d g/ are typically voiced, no matter what stress the following vowel has. However, for some speakers there is proportionately less voicing for /b/ (a surprising result) and/or before a syllable with main stress.

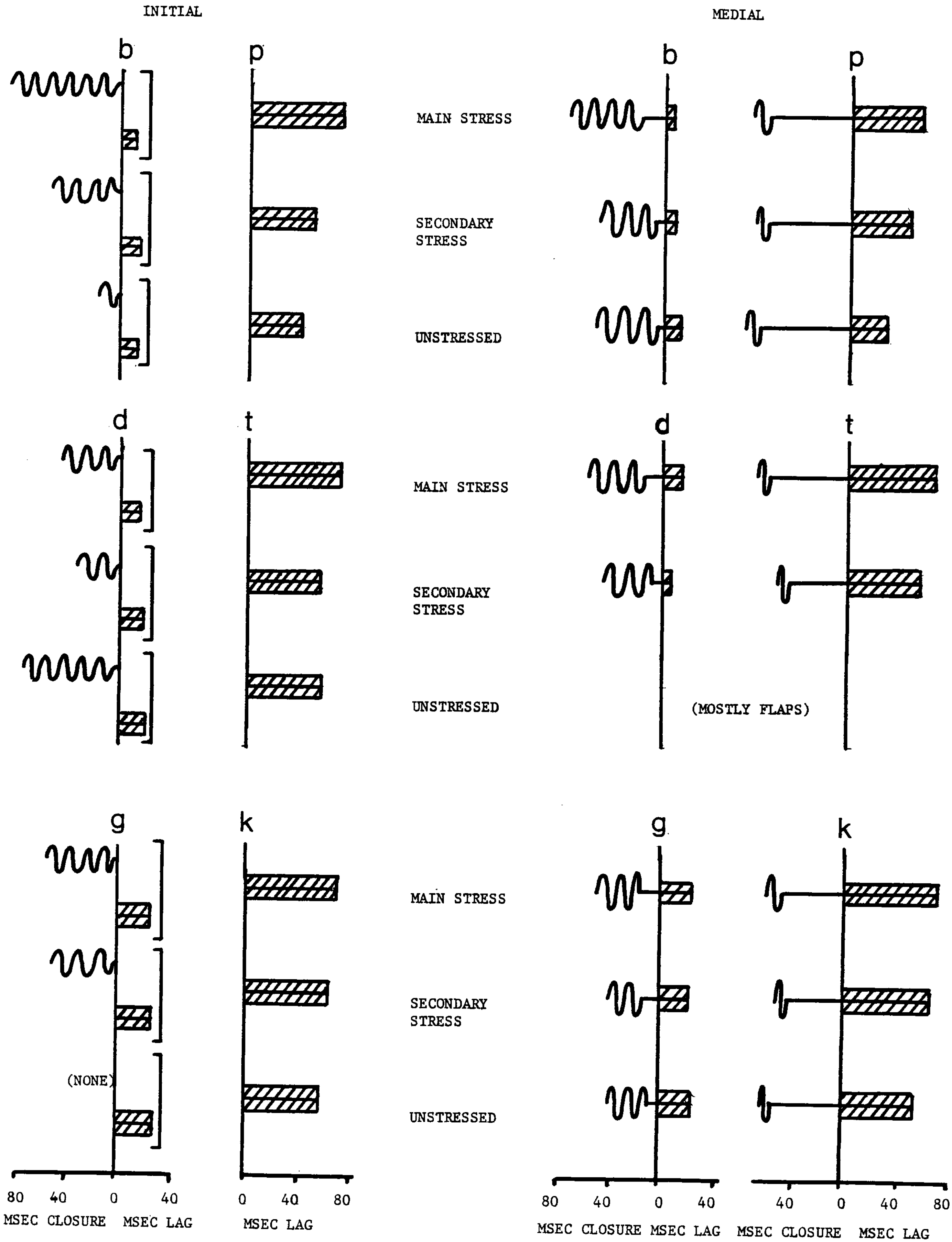


Figure 2. Mean acoustic measurements for English stops of six speakers.

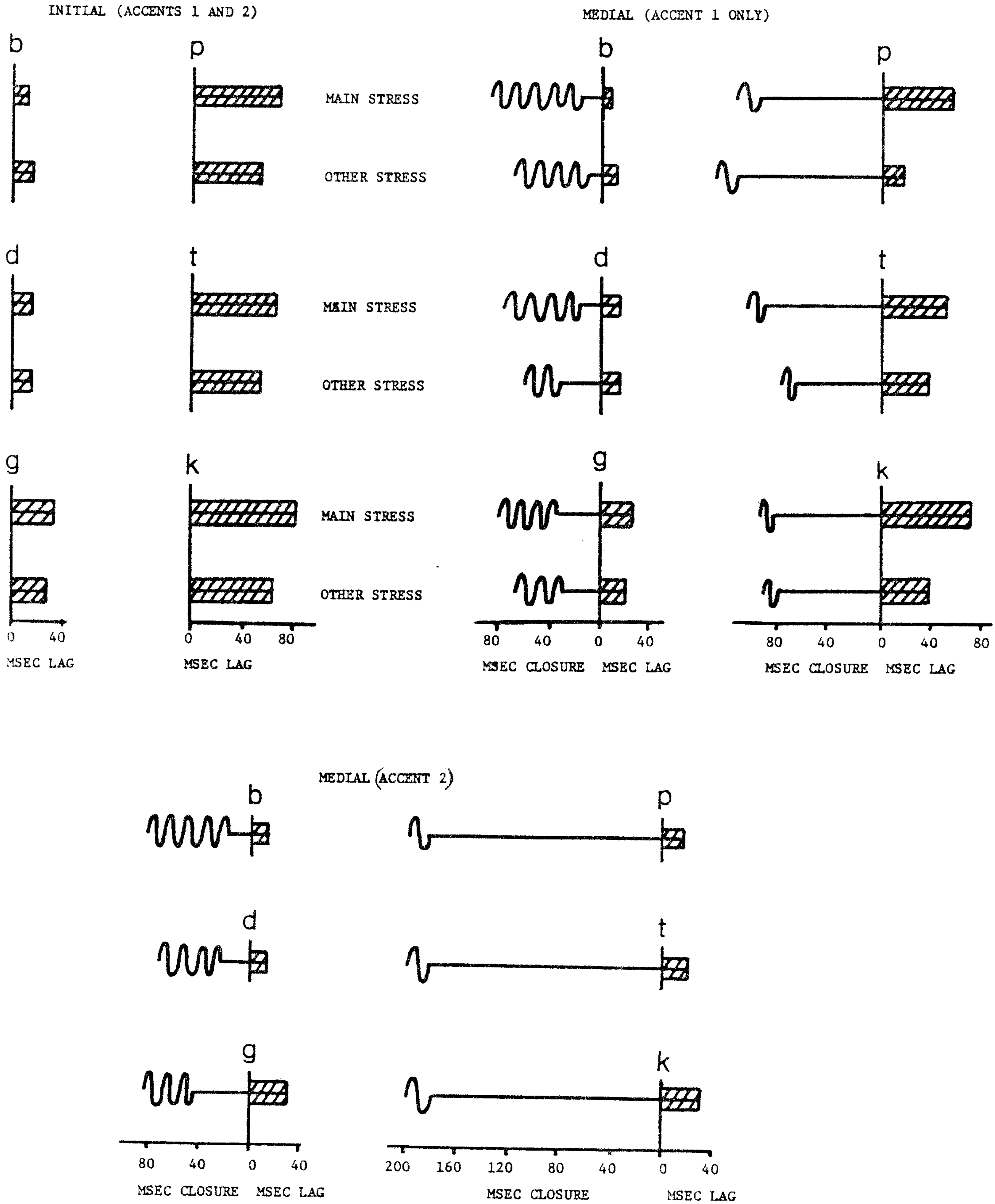


Figure 3. Mean acoustic measurements for Swedish stops of six speakers.

Results for Swedish are shown in Figure 3. The categories we tested were divided according to word accents, which require some explanation. A Swedish word has one of two pitch accents, which in the Stockholm dialect are markedly different from each other. Accent 1 words have one pitch peak, while Accent 2 words have two pitch peaks, one on each syllable. Only Accent 1 words can vary in where the stress falls; traditionally Accent 2 words are described as having primary stress on the first syllable and secondary stress on the second syllable. Therefore we separated Accent 1 from Accent 2 words in our list. Since our results show that as far as VOT is concerned, stops in Accent 2 words look like the corresponding unstressed stops in Accent 1 words, we have collapsed these categories for initial stops. The closure voicing of these categories is also the same. However, since stops in Accent 2 words have much longer voiceless intervals during their closures, we have not collapsed these categories in our presentation of the medial data. Medial Accent 2 data are shown separately in Figure 3, although in their VOT values they behave like unstressed Accent 1 stops.

To summarize the overall Swedish results, initial /b d g/ do not have closure voicing while medial /b d g/ do, with stress having only slight effects. Initial /p t k/ are aspirated, as are medial /p t k/ before main stress. However, other medial /p t k/ are generally unaspirated.

Our interest here has been grosser, categorial changes of the sort that might be noted in an impressionistic transcription and codified in a phonetic alphabet. However, various other effects of these variables on these measures have been the object of others' investigations (e.g. Zue 1976, Flege and Brown 1982). To some extent the patterns noted in those studies are seen here as well, but not always. However, we cannot expect to see effects on closure measures clearly, since the vowel preceding medial stops in our lists was not controlled. Nonetheless, we note briefly the effects evident in our data. For English, we see that stress increases the VOT for /p t k/; that place of articulation affects VOT, particularly for /b d g/. Place of articulation also affects the duration of closure voicing and the total closure duration; to a lesser extent the voicelessness measures, closure voicelessness and total voicelessness. Stress affects the duration of closure voicelessness for /b d g/, and of total voicelessness for all stops. For Swedish, the VOT measures are affected as for English. Also as for English, place of articulation affects the duration of closure voicing and for /b d g/ of total voicelessness. Stress affects the duration of closure voicing, for /b d g/ total closure, and for /p t k/ total voicelessness.

Discussion

To summarize, there is a clear overall trend in these languages to maintain aspirated /p t k/ before more stressed vowels, and to voice medial /b d g/. At the same time, there are differences between these two quite similar languages. While prevoicing is seen in English, it is at best rare in Swedish. There is also more variation across speakers in English regarding amount of medial voicing. In addition, there are differences with regard to closure durations, especially as conditioned by the pitch accents in Swedish.

Our language survey suggested a generalization that languages with initial 'aspiration' contrasts account for most of the allophonic differences seen between initial and medial stops. This variation appeared to be correlated with stress. In the two languages considered here in some detail, there is remarkable agreement in how stops vary across contexts, with position determining /b d g/ allophones, and position + stress determining /p t k/ allophones. However, there are also subtle differences between the two languages, to the extent that English speakers show more overall variation than Swedish speakers do. Such differences are part of speakers' knowledge of their language, and must be represented somewhere in their grammars. While most of the burden of the allophonic variation can possibly be born by some kind of general principle, at least some of the detail must be specified language by language. That is, the very lowest level of the grammar must contain language-specific quantitative rules.

Of course, this is not to say that all the detailed differences seen in our data must be represented in the phonology. To the extent that we can derive the small-scale effects of place of articulation and stress from physiology, they would not have to be stated for each language. That is, we can use cross-linguistic comparisons to determine which aspects are beyond language-specific control. In further work we can then attempt to account for the more general features of phonetic implementation, e.g. with aerodynamic modeling. This paper is intended as a first step in the direction of teasing apart the contributions of the phonetic and phonological components to the observed acoustic structure of voiced and voiceless stops.

Acknowledgements

This work was supported by UCLA and by an NSF grant to Peter Ladefoged. Earlier versions of this paper were presented at the 1982 Annual meeting of the LSA in San Diego and at the Conference on Causality and Language Change, UCLA, May 1982. The latter appears in UCLA WPP 54. We thank Don Brusasco and Ellen Jackson for help in data gathering and analysis, Jane Simpson for Australian language data sources, and Kathy Houlihan, Peter Ladefoged, and Ian Maddieson for comments.

Footnote

1. UPSID does contain a few examples of languages with voiced but no voiceless stops, but they were not included in our sample. Such descriptions are open to reinterpretation in light of the typical transcription of short lag VOT stops (voiceless unaspirated) as voiced, following English. The description of Yidiɲ makes such a reinterpretation likely but not certain.

References

Abdel-Massih, Ernest T. (1975). An Introduction to Egyptian Arabic. Center for Near Eastern and North African Studies, University of Michigan, Ann Arbor.

Allen, W. S. (1950). Notes on the phonetics of an Eastern Armenian speaker. Transactions of the Philological Society: 180-206.

Armstrong, Lilias E. (1932). The Phonetics of French. Twelfth printing, 1978, London.

Armstrong, Lilias E. (1940). The Phonetic and Tonal Structure of Kikuyu. Oxford University Press, London.

Aronson, H.I. (1968). Bulgarian Inflectional Morphophonology. Slavistic Printings and Reprintings 70, Mouton, The Hague.

Arun, Vidya Bhaskar (1961). A Comparative Phonology of Hindi and Panjabi. V. V. R. I. Press.

Azkue, R. M. (1949). El Vascuence. Bilbao.

Beach, D. M. (1938). The Phonetics of the Hottentot Language. Heffer Ltd., Cambridge.

Borgstrøm, Carl Hj. (1937). The Dialect of Barra in the Outer Hebrides. Norsk Tidssdrift for Sprogvidenskap VIII. Oslo.

Caramazza, Alfonso and Grace Yeni-Komshian (1974). Voice onset time in two French dialects. J. Phonetics 2: 239-45.

Chao, Yuen Ren (1947). Cantonese Primer. Publications for the Harvard Yenching Institute, Harvard University Press.

Chao, Yuen Ren (1968). A Grammar of Spoken Chinese. University of California Press, Berkeley.

Cole, Desmond T. (1955). An Introduction to Tswana Grammar. Longmans, Green and Co., London.

Daniels, F. J. (1958). The Sound System of Standard Japanese. Kenkyusha Ltd.

Dasgupta, S. (1971). Grundzüge einer Laut- und Schriftlehre des Bengalischen. Julius Groos Verlag Heidelberg.

Dent, Laurel. (1976). Voice Onset Time of Spontaneously Spoken Spanish Voiceless Stops. Presented at the 91st meeting of the Acoustical Society of America, April, 1976.

Dixon, R. M. W. (1977). A Grammar of Yidiɲ. Cambridge University Press.

Dow, Francis D. M. (1972). An Outline of Mandarin Phonetics. Faculty of Asian Studies in Association with the Australian National University Press, Canberra.

Fischer-Jørgensen, Eli (1954). Acoustic Analysis of Stop Consonants. Miscellanea Phonetica II:42-59. Also in Readings in Acoustic Phonetics, Ilse Lehiste, ed., M. I. T. Press.

Flege, James E., and William Brown (1982). Effects of utterance position on English speech timing. Phonetica, in press.

Haas, Mary R. (1956). The Thai System of Writing. American Council of Learned Societies, Washington, D. C.

Han, Mieko Shimizu (1962). Japanese Phonology: An Analysis Based Upon Sound Spectrograms. Kenkyusha Ltd.

Hardie, D. W. F. (1948). A Handbook of Modern Breton. University of Wales Press, Cardiff.

Harris, James W. (1969). Spanish Phonology. MIT Press.

Hashimoto, Oi-Kan Yue (1972). Phonology of Cantonese. (Studies in Yue Dialects 1). Princeton University Press, Princeton.

Hebert, R. J., and N. Poppe (1963). Kirghiz Manual. IUPUAS 33.

Holmer, Nils (1947). Critical and Comparative Grammar of the Cuna Language. Göteberg.

Houlihan, K. (1982). Is Intervocalic Voicing a Natural Rule? Paper presented at Annual meeting of LSA, San Diego.

Houlihan, K., and G. Iverson (1979). Functionally-constrained phonology. In D. A. Dinnsen (ed.), Current Approaches to Phonological Theory. Indiana University Press.

Jha, Aparna (1977). An Outline of Marathi Phonetics. Deccan College, Poona.

Jones, Daniel, and Dennis Ward (1969). The Phonetics of Russian. Cambridge University Press.

Kålmån, Bėla (1972). Hungarian Historical Phonology. Translated by Imre Gombos, in The Hungarian Language, ed. Lorånd Benkô and Samu Imre. Budapest.

Kahn, Daniel (1976). Syllable-Based Generalizations in English Phonology. MIT Dissertation, distributed by Indiana University Linguistics Club.

Kao, Diana L. (1971). Structure of the Syllable in Cantonese. Mouton, The Hague.

Kaufman, Terence (1971). Tzeltal Phonology and Morphology. University of California Press.

Keating, Patricia A. (1983). Phonetic and Phonological Representation of Stop Consonant Voicing. UCLA WPP, this volume.

Keating, Patricia A., Michael J. Mikoś, and William F. Ganong, III (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. J. Acoust. Soc. Am. 70:1261-71.

Kostić, Djordje, Alokananda Mitter, and K. C. Rastogi (1975). A Short Outline of Hindi Phonetics. Eka Press.

Kostić, Djordje, and Rhea S. Das (1972). A Short Outline of Bengali Phonetics. Calcutta.

Kraft, C. H. and M. G. Kraft (1973). Introductory Hausa. University of California Press, Berkeley.

Lisker, Leigh, and Arthur S. Abramson (1964). A cross-language study of voicing in initial stops: Acoustical measurements. Word 20: 384-422.

Lisker, Leigh, and Arthur S. Abramson (1967). Some effects of context on voice onset time in English stops. Lang. Speech 10:1-28.

Lundström, Eva, and Peter af Trampe. Elementär fonetik. Akademilitteratur, Stockholm.

Maddieson, I. (1981). UPSID: Data and Index. UCLA WPP 53.

Martin, Samuel E., and Young-Soo C. Lee (1969). Beginning Korean. Yale University Press.

Meyers, Laura (1976). Aspects of Hausa Tone. UCLA WPP 32.

Mikoś, Michael J. (1977). Problems in Polish Phonology. Unpublished dissertation, Brown University.

Mitchell, T. F. (1956). An Introduction to Egyptian Colloquial Arabic. Oxford University Press.

Morse, Mary Lynn (1980). A Sketch of the Phonology and Morphology of Bobo. University Microfilms International, Ann Arbor.

Obolensky, S., K. Y. Panah, and F. K. Nouri (1963). Persian Basic Course. Center for Applied Linguistics, Washington D. C.

Okell, John (1969). A Reference Grammar of Colloquial Burmese, Part I. Oxford University Press.

Osborne, C. R. (1974). The Tiwi Language. Australian Institute of Aboriginal Studies, Canberra.

Polomé, Edgar C. (1967). Swahili Language Handbook, Center for Applied Linguistics, Washington D. C.

Pukui, M. K., and S. H. Elbert (1965). Hawaiian-English Dictionary, 6th edition. University of Hawaii Press, Honolulu.

Redden, James E. (1979). A Descriptive Grammar of Ewondo. Occasional Papers on Linguistics No. 4, Southern Illinois University, Carbondale.

Schachter, P., and F. T. Otanes (1972). Tagalog Reference Grammar. University of California Press, Berkeley.

Shetter, William Z. (1961). Introduction to Dutch - A Practical Grammar. Martinus Nijhoff, The Hague.

Shimizu, Katsumasa (1979). Articulatory Effects on Speech Perception: Adaptation Test of Voicing Feature Detectors. Studia Phonologica XIII: 1-7.

Spore, Palle (1965). La Langue Danoise. Akademisk Forlag.

Suomi, Kari (1980). Voicing in English and Finnish Stops. Publications of the Department of Finnish and General Linguistics of the University of Turku.

Ternes, E. (1970). Grammaire Structurale du Breton de L'île de Groix. Winter, Heidelberg.

Thompson, Laurence C. (1965). A Vietnamese Grammar. University of Washington Press, Seattle.

Tovar, Antonio (1950). La Lengua Vasca. Icharopena: San Sebastian.

Valdman, Albert (1976). Introduction to French Phonology and Morphology. Newbury House Publishers.

Vanvik, Arne (1972). A Phonetic-Phonemic Analysis of Standard Eastern Norwegian, Part I. Norwegian Journal of Linguistics 26.

Ward, Ida C. (1933?). The Phonetic and Tonal Structure of Efik. Heffer & Sons Ltd., Cambridge.

Welmers, W. E. (1946). A Descriptive Grammar of Fanti. Language Dissertation 39, Linguistic Society of America. Philadelphia.

Westbury, John R., and Patricia A. Keating (1980). A model of stop consonant voicing and a theory of markedness. Paper presented at Annual meeting of LSA, San Antonio.

Westbury, John R., and Patricia A. Keating (forthcoming). Unpublished manuscript, revised and expanded version of Westbury and Keating (1980).

Wonderly, William (1951). Zoque II: Phonemes and Morphophonemes. IJAL 17.

Yeni-Komshian, Grace, Alfonso Caramazza, and Malcolm S. Preston (1977). A study of voicing in Lebanese Arabic. J. Phonetics, 5: 35-48.

Zue, Victor W. (1976). Acoustic Characteristics of Stop Consonants: A Controlled Study. MIT dissertation, distributed by IULC.

APPENDIX - Positional allophones*

Language	Source(s)	Phonemes	Initial Allophones	Medial Allophones	Final Allophones	Comments
Akan	Welmers (1946)	/p/ - /b/	[pʰ] [pʲ] - [b] [bʲ]	presumably same	none	affrication, palatalization generally before /i/,/e/
		/t/ - /d/	[tʰ] [ts] - [d] [dz]	presumably same	none	
		/k/ - /g/	[kʰ] [tʃ] - [g] [dʒ]	presumably same	none	
		/kw/ - /gw/	[kwʰ] [tʃwʲ] - [gw] [dʒwʲ]	presumably same	none	
Alyawarra	field notes, Nancy Turtle and Carol Morris	/p/	[p]	same	none	informal acoustic data, 1 speaker
		/t̪/	[t̪]	same	none	
		/t/	[t]	same	none	
		/ʈ/	[ʈ]	same	none	
		/tj/ (alveopalatal)	[tʃ]	same	none	
		/k/	[k]	same	none	
Arabic (Egyptian)	Abdel-Massih (1975) Mitchell (1956)	/b/	[b]	same	same	[p] in loans only
		/t/ - /d/	[t] - [d]	same	same	
		/ṭ/ - /ḍ/	[ṭ] - [ḍ]	same	same	
		/k/ - /g/	[k] - [g]	same	same	
		/ʔ/	[ʔ]	same	same	
Armenian (Eastern)	Allen (1950)	/pʰ/ - /p'/ - /b/	[pʰ] - [p'] - [b]	[pʰ] - [p'] - [b]	[pʰ] - [p'] [pˀ] - [b̥]	[b d g] after nasals
		/tʰ/ - /t'/ - /d/	[tʰ] - [t'] - [d]	[tʰ] - [t'] - [d]	[tʰ] - [t'] [tˀ] - [d̥]	initial [b d g] followed by low pitched breathy V
		/kʰ/ - /k'/ - /g/	[k] - [k'] - [g]	[kʰ] - [k'] - [g]	[kʰ] - [k'] [kˀ] - [g̊]	final /b d g/ are partly devoiced
		also 2 affricate sets				
Basque	Azkue (1949) Tovar (1950)	/p/ - /b/	[p] - [b]	same	none	/p t k/ voiced after nasals
		/t/ - /d/	[t] - [d]	same	[t]	
		/k/ - /g/	[k] - [g]	same	[k]	
Bengali	Kostic et al. (1972) Dasgupta (1971)	/p/ - /pʰ/ - /b/ - /bʱ/	[p] - [pʰ] - [b] - [bʰ] [ᵐbʱ]	same	same or [p] - [b]	
		/t̪/ - /t̪ʰ/ - /d̪/ - /d̪ʱ/	[t̪] - [t̪ʰ] - [d̪] - [d̪ʰ] [ⁿd̪ʱ]	same	same or [t̪] - [d̪]	also /t̪:/, /d̪:/
		/ʈ/ - /ʈʰ/ - /ɖ/ - /ɖʱ/	[ʈ] - [ʈʰ] - [ɖ] - [ɖʰ] [ⁿɖʱ]	same	none	
		/k/ - /kʰ/ - /g/ - /gʱ/	[k] - [kʰ] - [g] - [gʰ] [ŋgʱ]	same	same or [k] - [g]	also /k:/, /g:/
Bobo	Morse (1980)	/p/ - /b/	[pʰ] - [b]	[pʰ] - [b] [β]	none	few borrowed geminates and clusters
		/t/ - /d/	[tʰ] - [d]	same	none	
		/k/ - /g/	[kʰ] - [g]	same	none	palatals before [i], [ɪ]
Breton	Hardie (1948) Ternes (1970)	/p/ - /b/	[pʰ] - [b]	same	same	velars fronted before non-low V's
		/t̪/ - /d̪/	[t̪ʰ] - [d̪]	same	same	less aspiration before sonorants
		/c/ - /ɟ/	[cʰ] - [ɟ]	same	same	less aspiration after fricatives
		/k/ - /g/	[kʰ] - [g]	same	same	no aspiration before fricatives
						unreleased before obstruents
Bulgarian	Aronson (1968)	/p/ - /b/	[p] - [b]	same	[p]	regressive voicing assimilation
		/t/ - /d/	[t] - [d]	same	[t]	informal acoustic data for 1 speaker
		/k/ - /g/	[k] - [g]	same	[k]	
		also palatalized stops				

*Contrasting categories are separated by a dash. The transcriptions used are changed from the sources' somewhat for the sake of consistency, especially for the allophones.

'Same' as an entry under medial or final allophones means same as material to the left, for that language, and not material above.

Burmese	Okell (1969)	/p/ - /p^h/ - /b/ /t/ - /t^h/ - /d/ /c/ - /c^h/ - /ɟ/ /k/ - /k^h/ - /g/ /ʔ/	[p] - [p^h] - [b] [t] - [t^h] - [d] [c] - [c^h] - [ɟ] [k] - [k^h] - [g] none	[p] [b] - [p^h] [b] - [b] [t] [d] - [t^h] [d] - [d] [c] [ɟ] - [c^h] [ɟ] - [ɟ] [k] [g] - [k^h] [g] - [g] none	none none none none [p] [t̪] [t] [c] [k] [ʔ]	dialects differ in medial voicing palatals may be affricates final /ʔ/ assimilates in phrases
Cantonese	Chao (1947) Kao (1971) Hashimoto (1972)	/p/ - /p^h/ /t/ - /t^h/ /k/ - /k^h/ /kw/ - /kw^h/	[p] - [p^h] [t] - [t^h] [k] - [k^h] [kw] - [kw^h]	same same same same	[p˺] [p^ʔ] [t˺] [t^ʔ] [k˺] [k^ʔ] none	some tone co-occurrence restrictions
Choctaw	unpub. data, Linker	/p/ - /b/ /t/ /k/ /ʔ/	[p^h] [p] - [b] [ɓ] [t^h] [t] [k^h] [k] none	[p^h] [p] - [b] [t^h] [t] [d] [k^h] [k] [g] none	[p˺] [p] [t˺] [t] [k˺] [k] [ʔ]	[b d g] before non-stress also medial geminates also [k^w]
Cuna	Holmer (1947)	/p/ - /pp/ /t/ - /tt/ /k/ - /kk/	[p] [b̥] [t] [d̥] [k] [g̊]	[b] - [p:] [d] - [t:] [g] - [k:]	[p˺] [t˺] [k˺]	
Danish	Fischer-Jørgensen (1954) Spore (1965)	/p/ - /b/ /t/ - /d/ /k/ - /g/	[p^h] - [b̥] [t^h] - [d̥] [k^h] - [g̊]	[p^h] [b] - [b] [v] [t^h] [d] - [d] [ð] [k^h] [g] - [g] [ɣ]	[b] [d] [ð̆] [g] [ɣ]	Fischer-Jørgensen's acoustic data final free variation in voicing spirantization contrast before [ə]
Dutch	Shetter (1961)	/p/ - /b/ /t̪/ - /d̪/ /k/	[p] - [b] [t̪] - [d̪] [k]	same same same	[p] [t̪] same	
Efik	Ward (1933?)	/p/ - /b/ /t/ - /d/ /k/ /k͡p/	[b] [t] - [d] [k] [k͡p]	[b] [β] - [b] [t] [ɾ] - [ɾ] [k] [ɣ] same	[p] [t^s] [k] none	medial [t] before stress only medial [k] before stress only
English	Kahn (1976) Lisker and Abramson (1967)	/p/ - /b/ /t/ - /d/ /k/ - /g/	[p^h] - [b] [b] [t^h] - [d] [d] [k^h] - [g] [g]	same or [p] - [b] same or [ɾ] same or [k] - [g]	[p] [p˺] [p^ʔ] - [b] [b̥] [b˺] [t] [t˺] [t^ʔ] - [d] [d̥] [d˺] [k] [k˺] [k^ʔ] - [g] [g̊] [g˺]	medials 'same' before stress
Ewondo	Redden (1979)	/b/ /t/ - /d/ /k/ - /g/ /k͡p/ - /g͡b/	[b] [t] - [d] [k] [k͡p] - [g͡b]	[b] [β] [t] - [d] [ɹ] [k] - [g] [ɣ] same	[b̥] [d̥] [ɾ̥] [g̊] none	finals partly devoice
Finnish	Suomi (1980)	/p/ /t̪/ - /d/ /k/	[p] [t̪] [k]	[p] [t̪] [d] [k]	none [t̪] none	also /p: t̪: k:/ /d/ not underlying; some dialects [d̥] [ɾ]
French	Armstrong (1978) Valdman (1976)	/p/ - /b/ /t̪/ - /d̪/ /k/ - /g/	[p] - [b] [t̪] - [d̪] [k] - [g]	same same same	same same same	finals always released regressive voicing assimilation
Gaelic (Scots)	Borgstrøm (1937)	/p/ - /b/ /t̪/ - /d̪/ /k/ - /g/ also preaspirated stops	[p^h] - [b̥] [^{m}b] [t̪^{h}] - [d̪̥] [^{n}d̪] [k^h] - [g̊] [ŋg]	[p] - [mb] [t̪] - [n̪d̪] [k] - [ŋg]	[p˺] [t̪˺] [k˺]	also palatalization some voicing of /p t k/ after nasals

German	Moulton (1962) Korlén (1979)	/p/ - /b/ /t/ - /d/ /k/ - /g/	[p^h] - [b] [t^h] - [d] [k^h] - [g̊]	[p^h] [p] - [b] [t^h] [t] - [d] [k^h] [k] - [g]	[p^h] [t^h] [k^h]	deaspiration variable before nonstress
Hausa	Kraft and Kraft (1973) Meyers (1976)	/b/ - /ɓ/ /t̪/ - /d̪/ - /ɗ/ /k/ - /k'/ - /g/ /ʔ/	[b] [p] [p^h] [w] [f] - [ɓ] [b̰] [t̪^{h}] - [d̪] [d̪̥] - [ɗ] [d̰] [k^h] - [k'] - [g] [g̊] [ʔ]	same same same same	none none none none	medials less voiced than initials glottalization decreases voicing also palatalized, labialized allophones of velars
Hawaiian	Pukui and Elbert (1965)	/p/ /k/	[p] [k] [t]	same same	none none	informal acoustic data, 1 speaker
Hindi-Urdu	Kostić et al. (1975) Arun (1961)	/p/ - /b/ - /p^h/ - /b^ɦ/ /t̪/ - /d̪/ - /t̪^{h}/ - /d̪^{ɦ}/ /ʈ/ - /ɖ/ - /ʈ^h/ - /ɖ^ɦ/ /k/ - /g/ - /k^h/ - /g^ɦ/ /q/ (from Urdu) also affricate series	[p] - [b] - [p^h] - [b^ɦ] [t̪] - [d̪] - [t̪^{h}] - [d̪^{ɦ}] [ʈ] - [ɖ] - [ʈ^h] - [ɖ^ɦ] [k] - [g] - [k^h] - [g^ɦ] [q]	same same same same same	[p] - [b] [p] - [p] [ɸ] - [b^ɦ] [t̪] - [d̪] [t̪] - [t̪] - [d̪^{ɦ}] [ʈ] - [ɖ] - [ʈ] - [ɖ^ɦ] [k] - [g] - [k] - [g^ɦ] same	deaspiration before non-stress final /b d̪/ if not devoiced may be slightly implosive or unreleased
Hungarian	Kálmán (1972)	/p/ - /b/ /t̪/ - /d̪/ /k/ - /g/	[p] - [b] [t̪] - [d̪] [k] - [g]	same same same	same same same	
Japanese	Daniels (1958) Han (1962)	/p/ - /b/ /t̪/ - /d̪/ /k/ - /g/	[p] [p^h] - [b] [b̥] [t̪] [t̪^{h}] - [d̪] [d̪̥] [k] [k^h] - [g] [g̊]	same same same	none none none	sources disagree on extent of aspiration
Kaititj	field notes, Harold Koch	/p/ /t̪/ /t/ /ʈ/ /tj/ (lamino-palatal) /k/ also prenasalized stops	[p] [t̪] [t] [ʈ] [cç] [k]	same same same same same same	none none none none none none	also labialized stops
Kikuyu	Armstrong (1940)	/^{m}b/ /t/ - /^{n}d/ /c/ - /^{n}j/ /k/ - / g/	[^{m}b] [t] - [^{n}d] [tʃ] - [^{n}dʒ] [k] - [ŋg]	same same same same	none none none none	
Kirghiz	Hebert and Poppe (1963)	/p/ - /b/ /t̪/ - /d̪/ /q/ - /G/	[p^h] - [b̥] [t̪^{h}] - [d̪̥] [k^h] [q^h] - [g̊] [G̊]	[p] - [b] [t] - [d] [k] [q] - [g] [G]	[p] - [b̥] [b] [t] - [d̥] [d] [k] [q] - [g̊] [g] [G̊] [G]	final devoicing after V: or heavy stress also palatalized allophones
Korean	Martin and Lee (1969)	/p/ - /p^h/ - /P/ /t/ - /t^h/ - /T/ /k/ - /k^h/ - /K/	[p] - [p^h] - [P] [t] - [t^h] - [T] [k] - [k^h] - [K]	[b] - [p^h] - [P] [d] - [t^h] - [T] [g] - [k^h] - [K]	[p˚] [t˚] [k˚]	[p] slightly aspirated, [p^h] heavily [P] unaspirated, 'tense'; similarly, other places of artic. [b d g] also after nasals, laterals
Mandarin	Chao (1968) Dow (1972)	/p/ - /p^h/ /t/ - /t^h/ /k/ - /k^h/ **also 3 affricate sets**	[p] - [p^h] [t] - [t^h] [k] - [k^h]	same or [b] - [p] same or [d] - [t] same or [g] - [k]	none none none	medial voicing and deaspiration in toneless syllables

Marathi	Jha (1977)	/p/ - /p^h/ /t̪/ - /t̪^{h}/ - /d̪/ - /d̪^{ɦ}/ /ṭ/ - /ṭ^h/ - /ḍ/ - /ḍ^ɦ/ /k/ - /k^h/ - /g/ - /g^ɦ/ also dental and palatal affricates	[p] - [p^h] [t̪] - [t̪^{h}] - [d̪] - [d̪^{ɦ}] [ṭ] - [ṭ^h] - [ḍ] - [ḍ^ɦ] [k] - [k^h] - [g] - [g^ɦ]	same same [ṭ] - [ṭ^h] - [ɽ^h] same	same same [ṭ] - [ṭ^h] - [ɽ^h] same	
Nama	Beach (1938)	/p/ /t̪/ /k/	[p] [t̪] [k] [c]	[p] none none	[p^ʕ] none none	/t/, /k/ voiced after rising tone [c] before front vowels
Norwegian	Vanvik (1972)	/p/ - /b/ /t̪/ - /d̪/ /ṭ/ - /ḍ/ /k/ - /g/	[p^h] - [b] [t̪^{h}] - [d̪] none [k^h] - [g]	[p^h] [p] - [b] [t̪^{h}] [t̪] - [d̪] [ṭ^h] [ṭ] - [ḍ] [k^h] [k] - [g]	[p] - [b] [t̪] - [d̪] [ṭ] - [ḍ] [k] - [g]	informal acoustic data, 1 speaker deaspiration before non-stress but overall variation unclear
Persian	Obolensky et al. (1963)	/p/ - /b/ /t̪/ - /d̪/ /k/ - /g/ /G/	[p^h] - [b̥] [b] [t̪^{h}] - [d̪̥] [d̪] [k^h] - [g̊] [g] [q] [G]	same same same [q][ɢ][ʁ]	same same same same	variation somewhat unclear also /d:/
Polish	Mikoś (1977)	/p/ - /b/ /t̪/ - /d̪/ /k/ - /g/ also palatalized stops	[p] [p^h] - [b] [t̪] [t̪^{h}] - [d̪] [k] [k^h] - [g]	same same same	[p] [b̥] [t̪] [d̪̥] [k] [g̊]	
Russian	Jones and Ward (1969)	/p/ - /b/ /t̪/ - /d̪/ /k/ - /g/ also palatalized stops	[p] - [b] [t̪] - [d̪] [k] - [g]	same same same	[p] [t̪] [k]	
Spanish	Harris (1969)	/p/ - /b/ /t̪/ - /d̪/ /k/ - /g/	[p] - [b] [t̪] - [d̪] [k] - [g]	[p] - [β] [t̪] - [ð] [k] - [ɣ]	[β̞] [ð̞] [ɣ̞]	
Swahili	Polomé (1967)	/p/ - /p^h/ - /ɓ/ /t/ - /t^h/ - /ɗ/ /c/ - /c^h/ - /ʄ/ /k/ - /k^h/ - /ɠ/	[p] - [p^h] - [ɓ] [t] - [t^h] - [ɗ] [t] - [tʃ^h] - [dʒ] [k] - [k^h] - [ɠ]	same same same same	rare rare rare rare	stops not implosives after nasals fewer aspirates medially before unstressed V's velars backed before back V's
Swedish	Lundström and af Trampe (1977)	/p/ - /b/ /t̪/ - /d̪/ /ṭ/ - /ḍ/ /k/ - /g/	[p^h] - [b] [t̪^{h}] - [d̪] none [k^h] - [g]	[p^h] [p] - [b] [t̪^{h}] [t̪] - [d̪] [ṭ^h] [ṭ] - [ḍ] [k^h] [k] - [g]	[p^h] - [b] [t̪^{h}] - [d̪] [ṭ^h] - [ḍ] [k^h] - [g]	also geminates deaspiration before non-stress
Tagalog	Schachter and Otanes (1972)	/p/ - /b/ /t̪/ - /d̪/ /k/ - /g/	[p] - [b] [t̪] - [d̪] [k] - [g]	same same same	[p'] - [b'] [t̪'] - [d̪'] [k'] - [g']	informal acoustic data, 1 speaker
Tamil	pers. comm., Sarah Dart UCLA	/p/ - /pp/ /t̪/ - /t̪t̪/ /ṭ/ - /ṭṭ/ /k/ - /kk/	[p] [t̪] none [k]	[b] [β] - [p:] [d̪] [ð] - [t̪:] [ḍ] - [ṭ:] same [g] [ɣ] - [k:]	none none none none	slight aspiration before back vowels

Thai	Haas (1956)	/p/ - /p^h/ - /b/ /t/ - /t^h/ - /d/ /c/ - /c^h/ /k/ - /k^h/ /ʔ/	[p] - [p^h] - [b] [t] - [t^h] - [d] [c] - [c^h] [k] - [k^h] [ʔ]	presumably same presumably same presumably same presumably same same	[b̚] [d̚] none [g̚] same	
Tikar	pers. comm., Ellen Jackson, UCLA	/p/ - /b/ - /ɓ/ /t/ - /d/ - /ɗ/ /k/ - /g/	[p] - [b] - [ɓ] [t] - [d] - [ɗ] [k] ——— [g]	[p] - [mb] [β] - [ɓ] [β] [t] - [nd] - [ɗ] [k] - [g]	[p̚] [b̥̚] none none	
Tiwi	Osborne (1974)	/p/ /t̪/ /t/ /k/	[p] [b̥] [p^h] [t̪] [d̪] [t̪^{h}] [t] [d] [t^h] [k] [g̥] [k^h]	same same same same	none none none none	no intervocalic voicing
Tswana	Cole (1955)	/p'/ - /p^h/ - /b/ /t'/ - /t^h/ /k'/ - /k^h/ /tl'/ - /tl^h/	[p'] [p] - [p^h] - [b] [β] [t'] [t] - [t^h] [k'] [k] [c] - [k^x] [k^h] [c^h] [tl'] [tl] - [tl^h]	same same same same	none none none none	labialization before back vowels [c^h] before [i], [c] before front V's, [k^x] before [-high] V's
Tzeltal	Kaufman (1971)	/p/ - /p'/ - /b/ /t̪/ - /t̪'/ - /d̪/ /k/ - /k'/ - /g/	[p] - [p'] - [b] [t̪] - [t̪'] - [d̪] [k] - [k'] - [g]	[p] - [p'] - [β] [t̪] - [t̪'] - [ð] [k] - [k'] - [ɣ]	[p^h] - [p'] - [β̥] [t̪^{h}] - [t̪'] - [ð̥] [k^h] - [k'] - [ɣ̥]	/d̪/, /g/ in Spanish loans only finals may have voiceless V
Vietnamese	Thompson (1965)	/b/ /t/ - /t^h/ - /d/ /t̪/ /c/ /k/	[b] [ɓ] [t] - [t^h] - [d] [ɗ] [t̪] [c] [k]	presumably same presumably same presumably same presumably same presumably same	[p°] [t°] none [c˺] [k°]	also final [k͡p°] /ɣ/ has [g] after nasals
Yidiɲ	Dixon (1977)	/b/ /d/ /dɟ/ (alv. + pal.) /g/	[b] [b̥] [d] [d̥] [dɟ] [dɟ̊] [g] [g̊]	[b] [d] [dɟ] [g]	none none none none	
Zoque	Wonderly (1951)	/p/ - /b/ /t̪/ - /d/ /t^j/ - /d^j/ /c/ - /j/ /k/ - /g/	[p] - [b] [t̪] - [d] [t^j] - [d^j] [tʃ] - [ndʒ] [k] - [g]	[p] - [β] [t̪] - [ð] [t^j] [tʃ] [k] - [ɣ]	[p^h] [t^h] none [tʃ^h] [k^h]	degree of aspiration varies with position

"An Acoustical Study of the Register Distinction in Mon"

Thomas Lee

Introduction

Many Mon-Khmer languages show a register distinction in their vowels. It is believed that this register opposition historically evolved from a voicing distinction in consonant initials (Huffman 1976, Gregerson 1976). At some point in the development of these Mon-Khmer languages, allophonic differences began to emerge in the vowels so that voiceless initial stops conditioned one register (the first register), whereas voiced initial stops conditioned another register (the second register). When the voicing contrast in the initials was lost, the hitherto allophonic register differences in the vowels became a phonemic distinction.

The phonetic nature of this distinction has been described for various Mon-Khmer languages. In Khmer, for example, the first register is said to have a "normal" or "head" voice quality, usually accompanied by relatively higher pitch. The second register, on the other hand, is characterized by a deep, rather breathy, or "sepulchral" voice, larynx lowering, and relatively lower pitch (Henderson 1952).

In Mon, the register distinction has been described as "quasi-tonal" (Shorto 1962), again with the first register showing a comparatively higher pitch. In addition, the first register is associated with "clear" voice quality and relative tenseness, while the second register has breathy voice and is pronounced with "a general laxness of the speech organs" (Shorto 1966). As regards vowel quality, Shorto reports that first register vowels are peripheral in comparison to second register vowels, which are relatively centralized. First register vowels in Mon have also been reported to be higher in vowel height (Huffman 1976).

In another Mon-Khmer language, Brou, as Phillips, Miller and Miller (1963) observe, the first register vowels are pronounced with rather tense vocal cords and often with slight faucalization. In contrast, the second register vowels are produced with relatively relaxed vocal cords, with a rather deep and muffled sound. With regard to vowel quality in Brou, these linguists note that (unlike the situation in Mon) first register vowels are lower in height than their second register counterparts. They also note that pitch is not involved at all in the register distinction in Brou, pairs which differ only in register being said on the same pitch.

As can be seen from the above descriptions, the register distinction in Mon-Khmer may entail differences in phonation type, and pitch, as well as vowel quality. But the parameters involved in the distinction vary from one language to another; and even for the same parameter (e.g. vowel quality), the register distinction may be realized differently depending on the language concerned. However, despite these variations, significant similarities can be perceived among the register contrasts in various Mon-Khmer languages. Attempts have been made to seek a unitary explanation for these similarities. In this regard, Gregerson (1976) represents perhaps the most comprehensive treatment. He speculates that the acoustic manifestations of the register distinction in

Mon-Khmer could be due to a more fundamental physiological parameter: tongue-root position, with the first register having a more retracted tongue root, and the second register a more advanced tongue root.

The descriptions and hypotheses offered by linguists such as Henderson, Shorto and Gregerson on the registers of Mon-Khmer have by and large been based on impressionistic observations. In spite of their insights into the register issue, little is yet known about the physical and physiological parameters underlying the register distinction. This gap in research has been noted by Henderson (1977) and Ladefoged (1976). So far the only acoustical study relevant to the register question in Mon-Khmer seems to have been Miller (1967), an investigation of the vowel system of Brou. In that study, he could not find any acoustic evidence to substantiate a phonation difference between the two registers. Instead, he discovered that the register distinction may be cued by the directions of the movements of the first two formants within the vowel. In Brou, the register opposition primarily finds expression in non-short vowels. Among the non-short vowel nuclei, the second register vowels typically show two clear steady states in both F1 and F2, with the first steady state higher than the second steady state. The first register vowel nuclei reflect essentially the reverse situation, with the first steady state lower than the second. Miller suggests that for Brou, the register distinction is partly based on these formant structure cues.

This paper reports further acoustical findings on a Mon-Khmer language. The focus of the present study is Mon, a language spoken in Burma and parts of Thailand. Mon, as mentioned earlier, exhibits a register distinction which purportedly involves pitch, phonation type, and vowel quality. The first register vowels are plain-voiced and have a relatively higher pitch, while the second register vowels are breathy and lower-pitched. The latter are also reported to be slightly centralized in one study (Shorto 1966) and more open in another (Huffman 1976). Diffloth (1981) also observes that in Mon, the phonation difference between the two registers is fundamental, occurring in all phonetic contexts, whereas the pitch difference is optional. Diffloth notes that the pitch difference is a typical feature of word-by-word elicitation, and will disappear in the normal flow of speech. It is the purpose of the present investigation to explore the acoustical parameters governing the register distinction in Mon, and the relative significance of these parameters. With acoustical data, descriptions such as Shorto's and Diffloth's can be evaluated objectively.

Materials

The data for our acoustical analysis consist of 17 word pairs, all monosyllables, each said by five speakers of Mon (see Table I).[1] Sixteen of these word pairs are minimal pairs contrasting only in register, the remaining one being a near-minimal pair. For 11 of these word pairs, each speaker said only one member of a pair at a time, followed immediately by another speaker. For each of the other six word pairs, however, each speaker said both members of a pair in immediate succession. Two of these six word pairs were repeated in reverse order to offset any possible rhythmic effects on the tokens. Out of the 17 word pairs, three were also said in a sentence frame. Altogether, there were 95 (5 x 19) pairs of tokens of words spoken in isolation and 15 (3 x 5) pairs of tokens spoken in sentence frames.

Table I. List of Mon Words. The symbol (..) under the vowel indicates the second register. Pairs 15 and 16 were said twice; the first time, the second register token preceded its first register counterpart. The second time, the reverse order was followed. Pairs 9, 14, and 15 were also said in sentence frames:

a)	/kɒ	?oa	_______	ɲɛ̤?/			(for /hloik/);
	give	me		a little			
b)	/kɒ	?oa	_______	mṳa/			(for /lo̤ik/ and pair 15);
	give	me		one			
c)	/hɒm		_______	mṳa	toa	plo̤n/	(for pair 14)
	say			one	time	again	

1 a)	/kləŋ/	"to come"	10 a)	/kleəŋ/	"to bind"
b)	/klə̤ŋ/	"boat"	b)	/kle̤əŋ/	"to carry on shoulder"
2 a)	/kle?/	"to be lost"	11 a)	/toa/	"hand"
b)	/kle̤?/	"short"	b)	/to̤a/	"to slap"
3 a)	/kleh/	"to wear over shoulder"	12 a)	/pua/	"religious festival"
b)	/kle̤h/	"to carve meat"	b)	/pṳa/	"as much as"
4 a)	/to/	"cotton"	13 a)	/pʰɔa/	"satiated"
b)	/to̤/	"center"	b)	/pʰɔ̤a/	"calamity"
5 a)	/chan/	"love"	14 a)	/ɓoa/	"bean"
b)	/cha̤n/	"meditation"	b)	/ɓo̤a/	"elder sister"
6 a)	/ɲan/	"noisy"	15 a)	/ŋ ɲiək/	"small parrot"
b)	/ɲa̤n/	"intelligence"	b)	/ŋ ɲi̤ək/	"tooth"
7 a)	/kloik/	"pig"	16 a)	/sai/	"to separate"
b)	/klo̤ik/	"waistcloth"	b)	/sa̤i/	"skinny"
8 a)	/kloin/	"wolf"	17 a)	/pai/	"reject"
b)	/klo̤in/	"long"	b)	/pa̤i/	"spouse of elder sister"
9 a)	/hloik/	"moss"			
b)	/lo̤ik/	"book"			

Wide and narrow band spectrograms were made of all the tokens to obtain measurements of vowel duration, formant frequencies, and fundamental frequency. Formant frequencies were measured on the wide-band spectrograms at two points in the vowel: at one-third and two-thirds of the vowel duration. Fundamental frequency measurements were based on the narrow-band spectrograms. In general, the tenth harmonic or the highest harmonic that showed up equally clearly in both members of the word pair was chosen for measurement. Lines perpendicular to the base were drawn to intersect the chosen harmonic at the beginning, middle and end points of the vowel. In addition, a measurement was made at the peak value of the harmonic if this did not coincide with one of these measurement points. The fundamental frequency was calculated at each of these four points.

To examine the distribution of energy in the spectra in both registers, wide and narrow band spectra were taken at one-third of the vowel duration for 40 of the simple vowel tokens in the data. The wide band spectra served as an aid in locating formants, whereas the narrow band spectra provided information on the amplitude of F0 and that of the higher harmonics. The respective differences in intensity between F0 and F1, F2, F3, F4 were measured for each of these 40 tokens. The difference between the F0 amplitude and the mean amplitude of the first four formants was also calculated for each token. The amplitude of a formant was taken to be that of the highest harmonic in that formant.

Results

Vowel Duration

The mean duration of vowels in the different registers is given in Table II. In general, the second register vowels were longer in duration than their first register counterparts. This relationship holds for approximately 72% of the token pairs spoken in isolation, and 66% of the sample pairs said in sentences. The mean duration was 330.8 msec for register-2 vowels in isolation and 302.2 msec for the corresponding register-1 vowels. A paired t-test shows that the difference is statistically highly significant ($p < .001$). Vowel durations for the words spoken in sentences are substantially shorter and the difference between vowels of different registers is not significant.

Table II. Vowel Duration. ** indicates highly significant difference $p < .001$.

	Mean: Vowels in Isolation	Mean: Vowels in Sentences
Register 1	330.8 msec	215.1 msec
Register 2	302.2 msec**	205.2 msec

Formant Frequencies

The mean formant frequencies of F1 and F2 at one-third and two-thirds of the vowel duration are shown in Table III. There is no significant difference between the two registers in these values. The word pairs were then subclassified into simple vowels and diphthongs to see if a difference could be observed within each subclassification. Neither the simple vowels nor the diphthongs revealed any uniform tendencies correlating with the register difference. Our data, therefore, do not confirm Huffman's claim that second register vowels are more open. If this were true, a significant difference in F1 would be expected.

Table III. Formant Frequencies (Hz)

		F1		F2	
		at 1/3 duration	at 2/3 duration	at 1/3 duration	at 2/3 duration
Mean (Vowels in Isolation)	Register 1	592	614	1353	1503
	Register 2	578	604	1339	1486
Mean (Vowels in Sentences)	Register 1	450	538	1362	1434
	Register 2	470	545	1296	1393

Although no overall difference in formant frequencies was found, certain of the word pairs contained obvious vowel quality differences. To examine these, the four pairs of simple vowels in the data, /a, a̤; e, e̤; ə, ə̤; o, o̤/ were plotted in an F1/F2 space with the aid of a computer, following the procedure described in Davis (1976) (see Figure 1). There were ten tokens for each of the vowels /e, e̤, a, a̤/ and five tokens for each of the vowels /ə, ə̤, o, o̤/. Two sets of F1 and F2 values were used for each vowel token, corresponding respectively to the two points of measurement in the vowel. For each cluster of vowel points, an ellipse with radii of two standard deviations is drawn along axes oriented along the principal components. Thus, the ellipses enclose approximately 95% of the scatter along each axis. Figure 1 shows that a major vowel quality differences exist between the two registers only in the case of /e, e̤/, where the vowel of the second register is higher than that of the first register. Elsewhere, the register difference is relatively small.

A paired t-test was carried out for each of the four simple vowel contrasts. As shown in Table IV, the F1 difference between the two registers is statistically highly significant at both points in the vowel for e/e̤ (reflecting a vowel height difference between the two registers), but is not significant for either a/a̤ or o/o̤. As for ə/ə̤, the F1 difference is probably significant only at two-thirds of the vowel duration. With respect to the F2 difference between the registers, it is not significant for either e/e̤ or a/a̤, but is probably significant only at two-thirds of the vowel duration for ə/ə̤ and o/o̤. These results indicate that as far as simple vowels are concerned, no consistent vowel quality differences could be found between the first and second registers of Mon. The one simple vowel pair that shows a significant register difference, i.e. e/e̤, reveals that the second register vowel is actually higher than its first register counterpart, rather than lower. Neither Huffman (1976) nor Shorto (1966) predict the alignment of the vowels seen in these results.

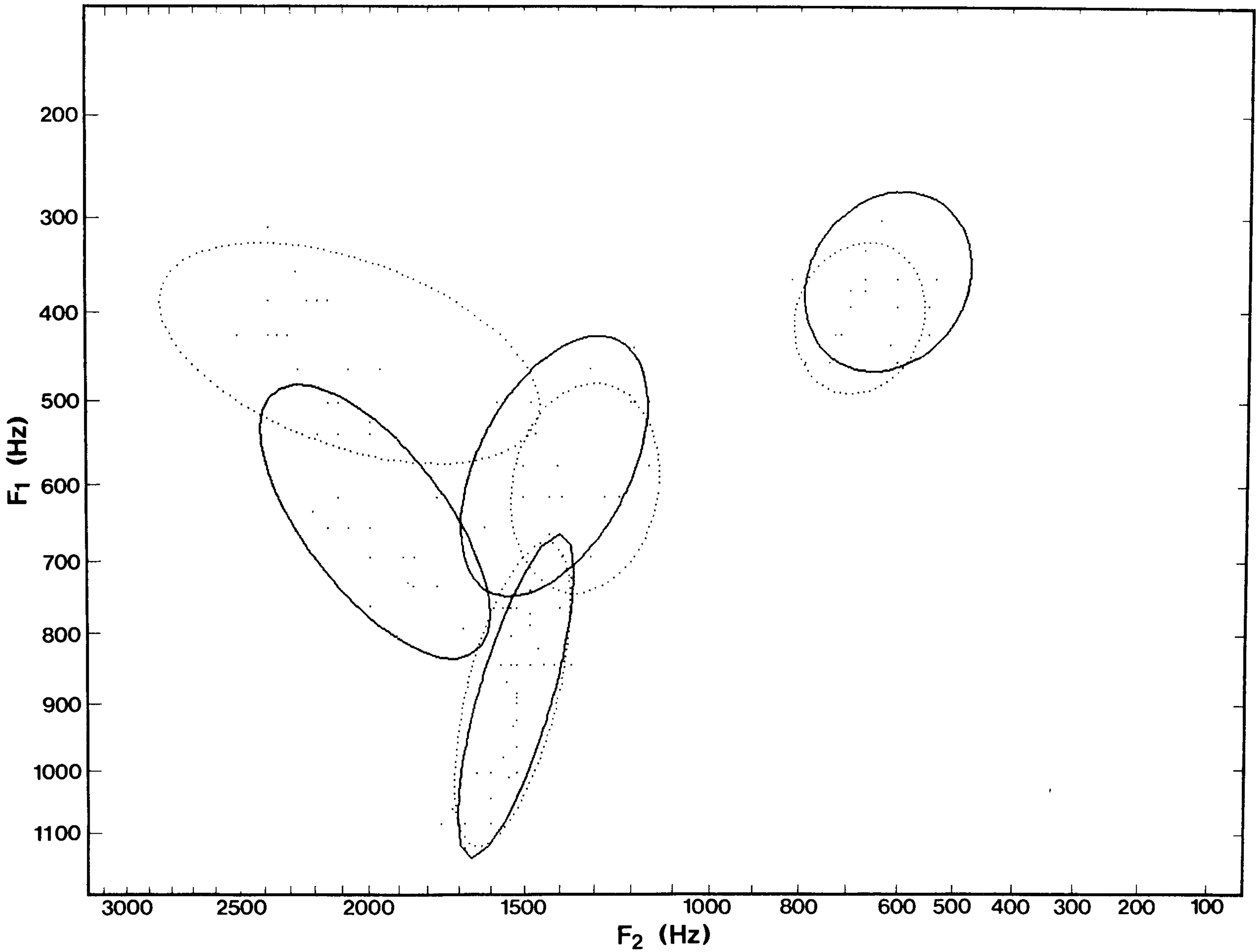


Figure 1. Simple Vowels of Five Mon Speakers. Solid lines indicate the first register, dotted lines the second register.

Table IV. Formant Frequencies of Simple Vowels in Isolation (Hz). ** indicates highly significant difference $p < .001$; *? indicates probable significant difference $p < .05$.

		F1		F2	
		at 1/3 duration	at 2/3 duration	at 1/3 duration	at 2/3 duration
Mean	Register 1	694	600	1840	2130
(for e/e̤)		**	**		
	Register 2	487	401	1965	2209
Mean	Register 1	881	894	1526	1519
(for a/a̤)					
	Register 2	879	890	1538	1528
Mean	Register 1	607	549	1385	1438
(for ə/ə̤)			*?		*?
	Register 2	631	578	1295	1351
Mean	Register 1	367	366	673	577
(for o/o̤)					*?
	Register 2	421	390	699	658

To see whether the second-register diphthongs of Mon are more centralized than the corresponding first-register diphthongs, as claimed by Shorto (1966), the diphthongs of each of the five Mon speakers were plotted. There were 15 tokens for each of the diphthongs /ai, a̤i, oi, o̤i/, ten tokens for each of the diphthongs /iə, i̤ə, oa, o̤a/, and five tokens for each of the following: /eə, e̤ə, ua, ṳa, ɔa, ɔ̤a/. As with the simple vowels, formant measurements were taken at one-third and two-thirds of the duration of the diphthong. A trajectory for each diphthong was then plotted using the means of these measurements.[2] For each speaker, the horizontal and vertical distances between the beginning and end points of a diphthong and the overall mean F1 and F2 values of that speaker (indicated by small triangles in the figures) were taken to be a rough measure of the degree of centralization of that vowel. A diphthong is considered more centralized than another diphthong if both the horizontal and vertical distance of its beginning and end points from the small triangle are shorter than the corresponding values of the other diphthong. On the whole, the diphthong plots show that there is no consistent centralizing tendency among the second-register diphthongs. Figures 2 and 3, representing the diphthongs of Speaker 2 and Speaker 4 respectively, illustrate this lack of consistency in centralization.

In summary, our findings suggest that, contrary to what has been reported, no consistent, <u>overall</u> vowel quality differences exist between the two registers of Mon. But there are differences in individual pairs of vowels, and considerable individual differences.

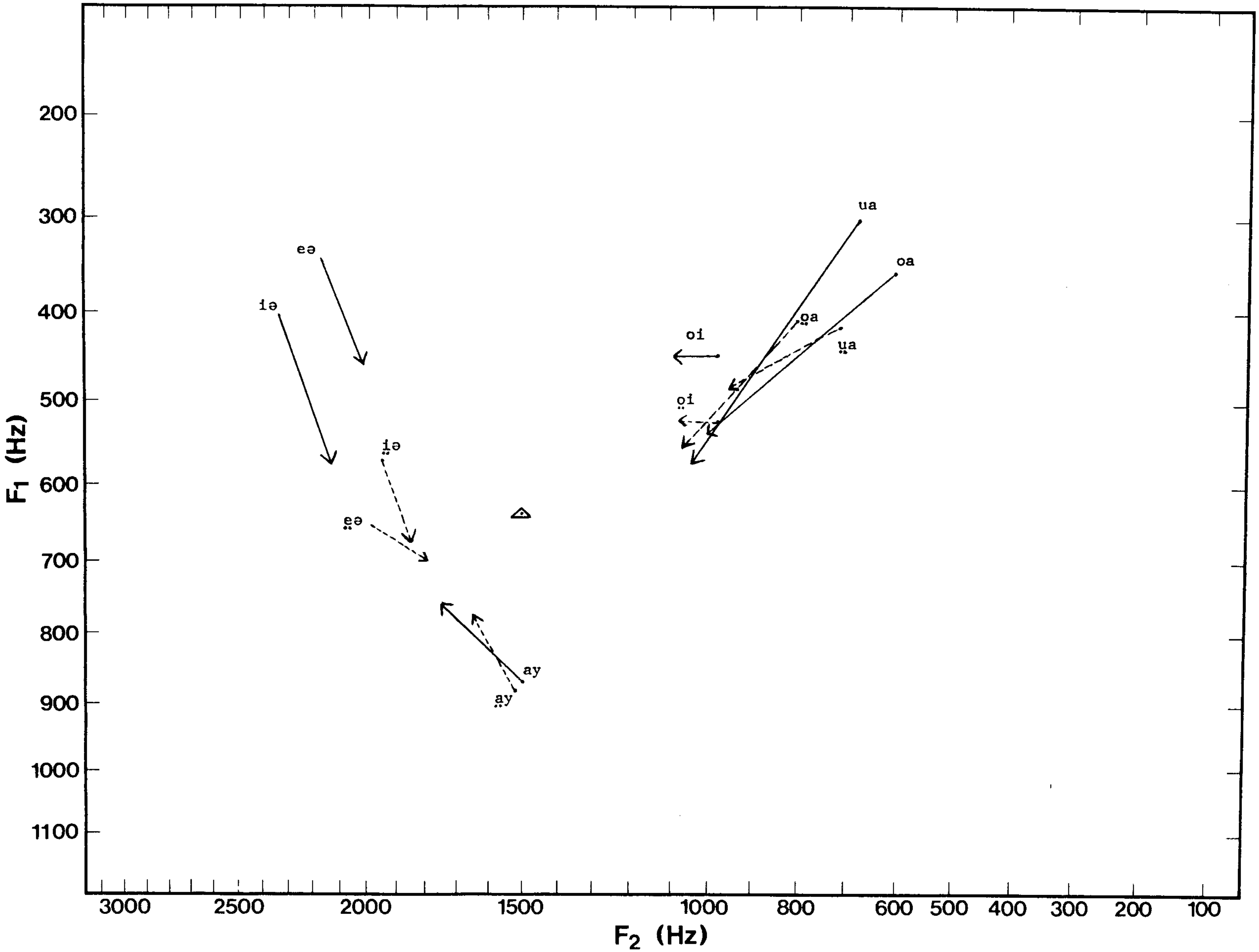


Figure 2. Diphthongs of Speaker 2. Solid lines indicate the first register, dash lines the second register.

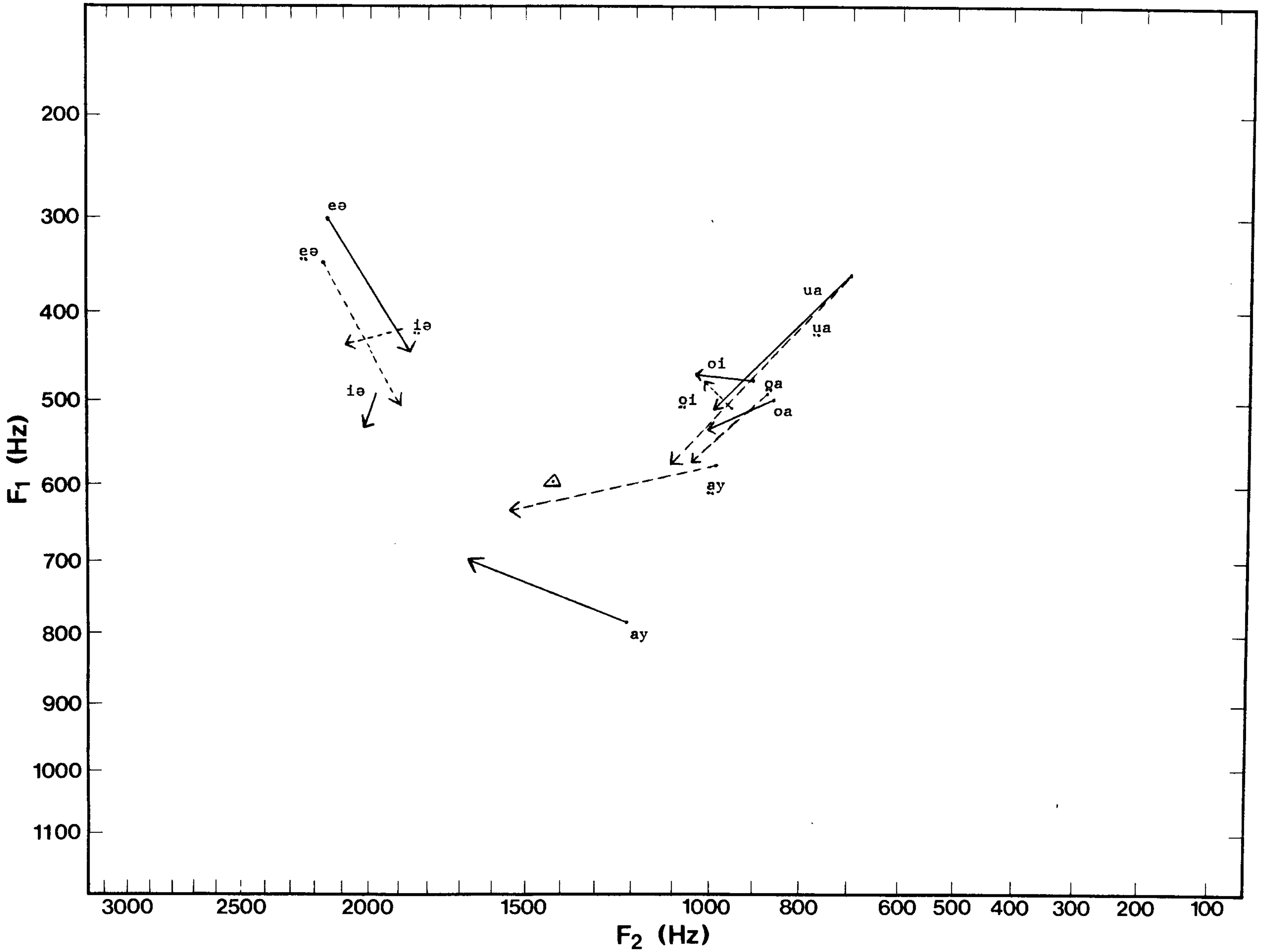


Figure 3. Diphthongs of Speaker 4. Solid lines indicate the first register, dash lines the second register.

Fundamental Frequency

The register distinction in Mon was most clearly reflected in the fundamental frequency of the vowels. The onset F0, mid-point F0, final F0, and maximum F0 of each register-1 token were paired with the corresponding values of its register-2 counterpart, and a paired t-test was carried out. The results (given in Table V) show that the difference in onset, mid and maximum F0 between the citation forms of the two registers was highly significant in each case, the difference being most prominent in onset F0. The register difference in final F0 was also significant. As for vowels in sentence frames, the differences in onset, mid and final F0 between the two registers were also significant, though the difference was not as great as that for vowels in isolation.

Table V. Fundamental Frequency in Mon Vowels (Hz). ** indicates highly significant difference, $p < .001$; * indicates significant difference, $p < .01$.

		Onset F_o	Mid F_o	Final F_o	Maximum F_o
Mean	Register 1	179	182	155	190
(Vowels in		**	**	*	**
Isolation)	Register 2	148	158	149	164
Mean	Register 1	169	171	158	174
(Vowels in		**	**	*	**
Sentences)	Register 2	136	144	146	150

The statistics in Table V demonstrate convincingly that at the onset, midpoint, and endpoint of the harmonics, first-register vowels show a higher pitch. Here, two questions could be raised. One is whether the general pitch level of first register vowels is higher than that of second register vowels. Another question concerns the role of pitch contour: to what extent does pitch contour reflect the register distinction?

To answer the first question, the onset, mid, and final F0 values of all the first-register tokens of each speaker were compared as a whole to the corresponding set of values in the second register. This comparison was made for each speaker by means of a grouped t-test, both for vowels in isolation and vowels in sentence frames. The mean F0 values of each speaker are given in Table VI. With regard to citation forms of vowels, the findings in the table point to a highly significant global pitch-level difference between the two registers for each of the five speakers ($p< .001$). As for vowels in sentence frames, a significant difference in overall pitch level was also observed between the two registers, but only for four of the five Mon speakers.

Table VI. Fundamental Frequency Level of Mon Speakers (Hz). ** indicates highly significant differences, $p < .001$; * indicates significant difference, $p < .01$.

		Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5
Mean F_o (Vowels in Isolation)	Register 1	166	188	179	166	158
		**	**	**	**	**
	Register 2	144	175	146	143	148
Mean F_o (Vowels in Sentences)	Register 1	161	198	166	161	144
		*	*	*	*	
	Register 2	136	171	133	129	142

The possible role of pitch contour in the register distinction of Mon was studied by comparing the F0 values of the two registers at three points along the time axis: the beginning, mid-point, and end-point of the harmonics. Figure 4 illustrates the F0 contour of each speaker for vowels in isolation; figure 5 illustrates the F0 contour of the speakers for vowels in sentences. It is clear from the figures that the register distinction is not expressed by a constant contour difference. For the sample taken as a whole, the pitch contour is in fact fairly random. Table VII gives the coefficients of variation of the differences between mid F0 and onset F0 and of the differences between final F0 and mid F0 for both registers. The randomness in pitch contour is reflected by the large coefficients of variation reported in the table.

Table VII. Fundamental Frequency Differences Within Registers (Hz). Coefficient of Variation (C.V.) = Standard Deviation/Mean x 100.

		Difference between Mid F_o and Onset F_o			Difference between Final F_o and Mid F_o		
		Mean	S.D.	C.V.	Mean	S.D.	C.V.
Vowels in Isolation	Register 1	2.8 Hz	14.7 Hz	525.0	-26.0 Hz	24.0 Hz	- 92.3
	Register 2	10.0 Hz	13.5 Hz	135.0	- 9.0 Hz	22.5 Hz	-250.0
Vowels in Sentences	Register 1	2.2 Hz	5.6 Hz	254.5	-13.1 Hz	12.8 Hz	- 97.7
	Register 2	7.3 Hz	10.6 Hz	145.2	2.2 Hz	8.2 Hz	373.0

It should be pointed out, however, that although there is no consistent pitch contour difference correlating with the register distinction, different speakers seem to use different but fairly consistent pitch patterns to mark the registers. Figures 4 and 5 illustrate this extremely interesting aspect of the pitch differences between the two Mon registers. Figure 4 indicates that in the first register, Speaker 1 and Speaker 4 show a rise-fall contour, with a slight pitch rise from the onset to the mid-point and a sharp fall from the mid-point to the end of the contour. Speakers 2 and 3, however, follow a gradually declining contour in the first register, while Speaker 5 uses an almost level pitch. With

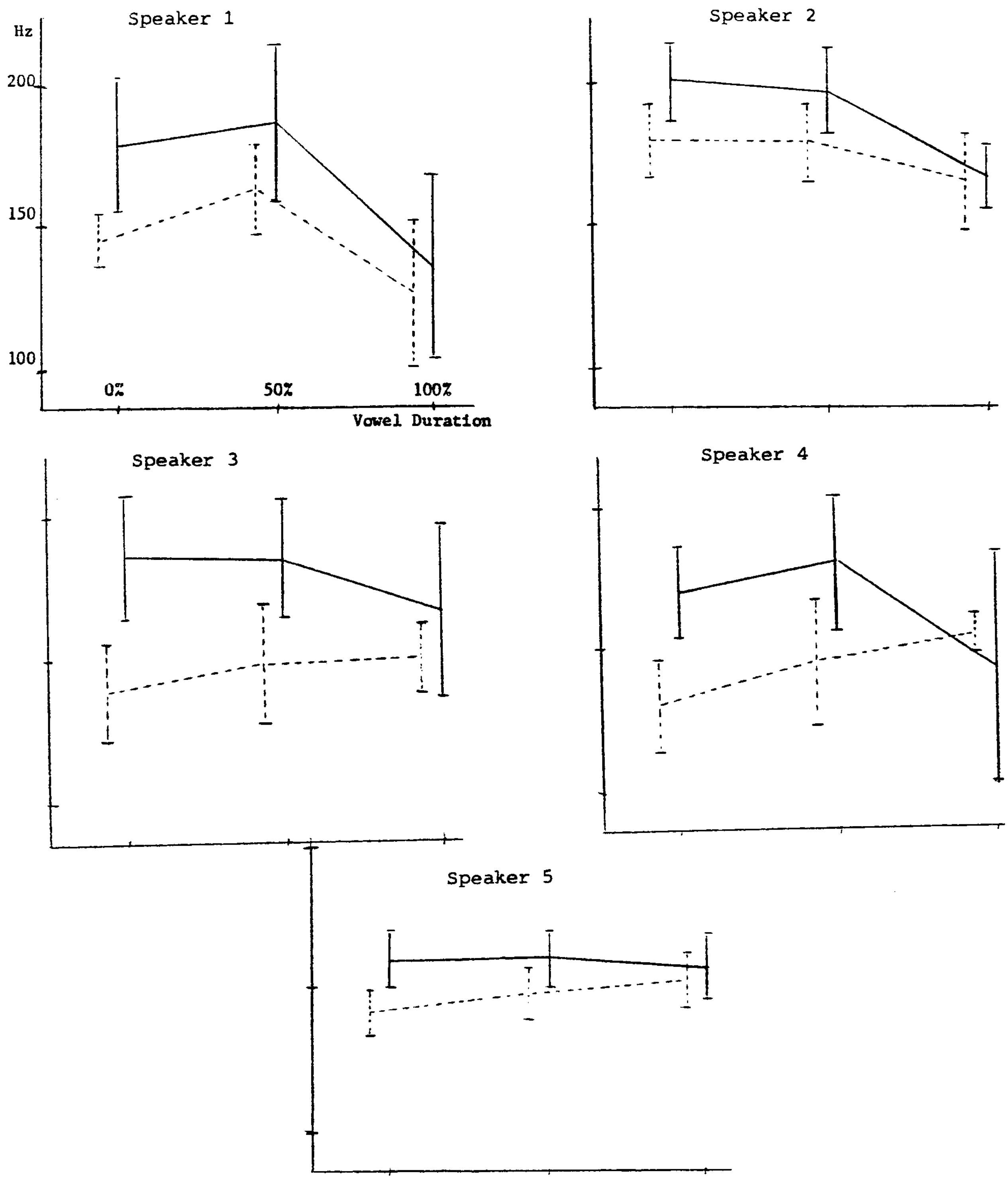


Figure 4. Pitch Contours of Five Mon Speakers for Vowels in Isolation. Solid lines indicate the first register, dash lines the second register. Each vertical bar represents a magnitude of two standard deviations.

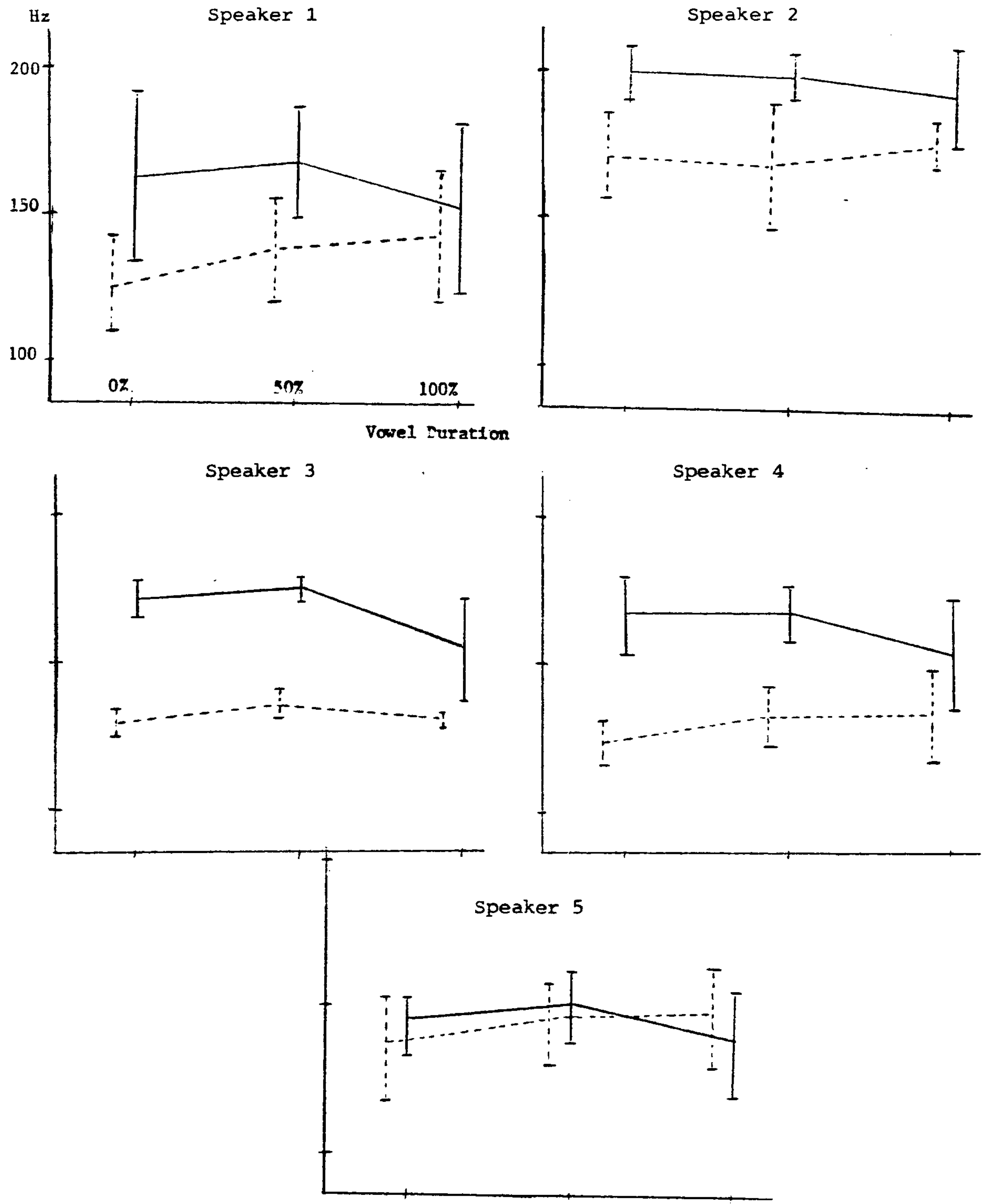


Figure 5. Pitch Contours of Five Mon Speakers for Vowels in Sentences. Solid lines indicate the first register, dash lines the second register. Each vertical bar represents a magnitude of two standard deviations.

respect to the second register, Speakers 3, 4 and 5 show a steadily rising pitch pattern, but Speaker 1 uses a rise-fall contour and Speaker 2 a slightly falling contour. Referring to Figure 5, which shows the pitch contours of vowels in sentences, one finds, however, that the contour differences are sharply reduced.

To find out whether for each speaker, the pitch contours of the two registers are significantly different, the difference between mid F0 and onset F0, and that between final F0 and mid F0, in the first register were paired with those of the second register. Table VIII and Table IX show respectively the mean values of these differences for vowels in isolation and vowels in sentences. In general, this difference was found to be statistically significant for all five speakers vis-a-vis vowels in isolation ($p<.025$). The absolute pitch level difference between the registers is relatively small for Speaker 5, compared to other speakers, but nonetheless the contour difference is statistically significant.

Table VIII. Pitch Contour Differences between Registers (Hz) for Vowels in Isolation. ** indicates highly significant difference, $p < .001$. * indicates significant difference, $p < .01$. *? indicates probable significant difference, $p < .05$.

		Mid F_o - Onset F_o	Final F_o - Mid F_o
Mean Difference (Speaker 1)	Register 1	8.4	-51.5
		**	**
	Register 2	18.7	-37.5
Mean Difference (Speaker 2)	Register 1	- 4.7	-30.3
			**
	Register 2	- 1.8	-13.8
Mean Difference (Speaker 3)	Register 1	- 1.4	-18.9
			*?
	Register 2	8.2	1.2
Mean Difference (Speaker 4)	Register 1	10.9	-30.3
		*?	*
	Register 2	19.1	- 0.58
Mean Difference (Speaker 5)	Register 1	0.9	- 3.4
		*?	*
	Register 2	5.8	5.5

Table IX. Pitch Contour Differences between Registers (Hz) for Vowels in Sentences. *? indicates probable significant difference, $p < .05$.

		Mid F_o - Onset F_o	Final F_o - Mid F_o
Mean Difference (Speaker 1)	Register 1	5.3	-15.0
	Register 2	14.0	- 4.7
Mean Difference (Speaker 2)	Register 1	- 1.7	- 5.7
	Register 2	- 3.0	9.7
Mean Difference (Speaker 3)	Register 1	2.3	-19.3
	Register 2	7.0	- 5.3
Mean Difference (Speaker 4)	Register 1	- 0.3	-14.0 *?
	Register 2	8.7	1.0
Mean Difference (Speaker 5)	Register 1	5.3	-11.7 *?
	Register 2	10.0	1.0

As regards vowels in sentences, as can be seen from Table IX, the pitch contour differences between the two registers were in general not significant. If one recalls Diffloth's (1981) remark that pitch differences in Mon disappear in the normal fow of speech, our findings indicate that what actually happens for vowels in sentences is that the pitch contour differences are diminished to only marginal significance. The overall pitch level differences, however, still remains, as reported in the preceding paragraphs.

Distribution of Spectral Engergy

It has been reported in the phonetic literature that phonation type differences are reflected in spectral energy distribution. For example, Stevens (1981) suggests that a relatively higher amplitude of F0 compared to that of the higher harmonics is typical of breathy phonation. In an earlier study, Fischer-Jorgenson (1967) discovered that Gujarati breathy vowels showed a stronger level of F0 amplitude in comparison with the amplitudes of F1, F2, and F4. More recently, Ladefoged (1982) found that in the Khoisan language !Xõo, the breathy vowels exhibited a higher amplitude of F0 relative to that of F1. To explore possible phonation differences between the two Mon registers, the energy distribution in the spectra of simple Mon vowels was examined.

The spectra were taken at one-third of the vowel duration, and the respective amplitude differences between F0 and F1, F2, F3, F4 as well as the difference between F0 amplitude and the mean amplitude of the first four formants, were measured. Since formant frequency differences will complicate any comparison of energy distribution in vowel spectra, only those simple vowels which do not show significant formant frequency differences between the registers at one-third of the vowel duration, i.e. /a, a̤, o, o̤, ə, ə̤/ were included in this part of the analysis (see Table IV).

There were ten tokens for the vowel pair a/a̤, and five tokens for each of the vowel pairs /ə, ə̤, o, o̤/. Again, the values obtained for the first register tokens were matched against corresponding values of the second register. A paired t-test was carried out for each of the significant relative amplitude difference between the two registers for a/a̤ and ə/ə̤. The vowel pair o/o̤, however, shows in the second register a stronger amplitude level of F0 relative to that of F1, F2, F3 and to the mean amplitude of the first four formants.

Table X. Relative Amplitude of F0 (dB) in relation to F1, F2, F3, F4. L_o represents the amplitude of the fundamental, L_n the amplitude of the nth formant. L_{av} is the average amplitude of the first four formants. * indicates significant difference, $p < .01$. *? indicates probable significant difference, $p < .05$.

		$L_o - L_1$	$L_o - L_2$	$L_o - L_3$	$L_o - L_4$	$L_o - L_{av}$
a/a̤	Register 1	-11.1	-17.0	- 5.5	- 3.6	- 9.5
	Register 2	-10.1	-15.7	- 3.1	- 1.2	- 7.5
ə/ə̤	Register 1	- 8.9	-14.3	- 3.4	2.0	- 6.9
	Register 2	-10.6	-12.7	- 1.4	4.2	- 6.5
o/o̤	Register 1	-12.0	- 7.7	5.3	3.3	- 2.8
		*?	*	*?		*
	Register 2	- 8.2	- 0.4	10.5	6.8	2.2

Our findings indicate that a phonation difference exists between /o/ and /o̤/, but no overall phonation difference can be found between the two registers. Our study, then, does not support Diffloth's claim that the fundamental difference between the two Mon registers lies in a phonation difference.

Concluding Remarks

In this investigation of Mon register, four parameters were looked at: vowel duration, frequencies of the first two formants, fundamental frequency, and distribution of spectral energy. It was found that for citation forms, significant differences exist between the two registers for two of these parameters. In contrast to the first register vowels, the second register vowels are characterized by (1) longer duration, and (2) a lower onset pitch and a lower overall pitch level. No consistent difference could be found between the two registers with respect to vowel quality or phonation. Pitch contour is also not a relevant parameter; speakers differ from one another in their pitch patterns; each speaker uses consistent but different contours to mark the two registers. As for vowels in sentence frames, the onset pitch level and the overall pitch level serve as good indicators of the register distinction.

Our findings confirm some of Shorto's and Diffloth's observations on Mon, to wit, their comments on the existence of pitch differences between the registers. Further, our findings suggest that the most significant parameter of the register distinction is that of pitch, in particular the onset F0 and the overall pitch level. Indeed, as Shorto (1962) suggests, Mon is a quasi-tonal language.

Notes

1. This word list was extracted from a larger corpus of recordings, which include many non-minimal pairs. The recording was made by Peter Ladefoged and Gerard Diffloth in Thailand in January 1981. I am indebted to them for allowing me to use their data.

2. The diphthongs /oa, ɔa/ were extremely close to each other in the vowel chart, and so were /o̤a, ɔ̤a/. For clarity of illustration, the tokens of /ɔa/ were grouped together with those of /oa/ and the tokens of /ɔ̤a/ with those of /o̤a/.

Acknowledgments

I am most grateful to Peter Ladefoged and Ian Maddieson for their guidance and encouragement in this project. I would also like to thank Pat Keating and Helen Tsai for their help on questions of statistics. This work was supported by NIH.

References

Davis, S. B. (1976). "Computer Evaluation of Laryngeal Pathology Based on Inverse Filtering of Speech." SCRL Monograph, No. 13. Speech Communication Research Laboratory: Santa Barbara, CA.

Diffloth, Gerard (1981). "Proto-Mon Registers: Two, Three, Four..?" Unpublished manuscript.

Fischer-Jorgensen, E. (1967). "Phonetic Analysis of Breathy (murmured) Vowels in Gujarati." Indian Linguistics 28.

Gregerson, Kenneth J. (1976). "Tongue-Root and Register in Mon-Khmer," in Austroasiatic Studies Part I. ed. by Philip Jenner, L. Thompson, and S. Starosta. Honolulu: University of Hawaii Press.

Henderson, Eugenie J. A. (1952). "The Main Features of Cambodian Pronunciation." Bulletin of the School of Oriental and African Studies 14:149-173.

Henderson, Eugenie J. A. (1977). "The Larynx and Language: A Missing Dimension?" Phonetica 34:256-263.

Huffman, Franklin E. (1976). "The Register Problem in Fifteen Mon-Khmer Languages." in Austroasiatic Studies Part I.

Ladefoged, Peter (1976). "Some Experimental Data on Phonation Types." Unpublished manuscript.

Ladefoged, Peter (1982). "The Linguistic Use of Different Phonation Types." UCLA Working Papers in Phonetics 54.

Miller, J. D. (1967). "An Acoustical Study of Brou Vowels." Phonetica 17:149-177.

Phillips, R., Miller, J., and Miller C. (1963). "The Brou Vowel System: Alternate Analyses." Unpublished manuscript. Cited in J. Miller (1967).

Shorto, H. L. (1962). A Dictionary of Modern Spoken Mon. London: Oxford University Press.

Shorto, H. L. (1966). "Mon Vowel Systems: A Problem in Phonological Statement." in In Memory of J. R. Firth ed. by C. E. Bazell et al. London: Longmans.

Stevens, Kenneth N. (1981). "Vibration Modes in Relation to Model Parameters" in Vocal Fold Physiology ed. by K. Stevens and M. Hirano. Tokyo: University of Tokyo Press.

"The Vowel System in Two Varieties of Cantonese"

Thomas Lee

The purpose of this research project is to look into the differences between the vowels of Cantonese as spoken in the city of Canton (Guangzhou) and those of Cantonese as spoken in the city of Hong Kong. The two speech communities of Canton and Hong Kong have undergone different social and economic developments in the past thirty three years (since 1949). The languages spoken by the two communities have also been in contact with different linguistic systems. The Cantonese spoken in Canton has been in contact mainly with Mandarin Chinese, the official language of China, whereas Hong Kong Cantonese has been in contact with English, which is an official language of Hong Kong. It may be assumed that under these divergent social, economic and linguistic influences, the linguistic systems of these two varieties of Cantonese may reveal shifts in different directions. This assumption is supported by the fact that quite a number of native speakers of either Canton or Hong Kong Cantonese whom I've talked to have remarked on their perception of differences between the two varieties of the same dialect. As a native speaker of Hong Kong Cantonese, I feel the vowels of these two varieties differ somewhat in phonetic quality. The immediate aim of this project, then, is to test whether my perceptions will be borne out by acoustic findings. The recording of these differences presumably will be of value in later studies of sound change in this dialect.

The phonetic system of Cantonese consists of 13 vowels, as indicated below. Although scholars may vary slightly from one another in their assignment of labels to these vowel allophones, and may diverge considerably in their grouping of these vowels into vowel phonemes, they generally agree on the existence of 13 vowels. Seven of these vowels are long (or tense), and six of them are short (or lax).

```
i:      y:              u:
  I                     U
 e      ø               o
  ɛ:     oe:            ɔ:
             ɒ
               ɑ:
```

The distributions of these 13 vowels with respect to syllable endings are given in Table 1, which is based on Hashimoto (1972:90).

As can be seen from Table 1, while long vowels can occur in diphthongs, and in open as well as closed syllables, short vowels can only occur in diphthongs and closed syllables. With the exception of [e] and [o], which only appear in diphthongs, all the other vowels can be immediately followed by some nasal or stop consonant endings. In this project, only vowels that can occur in closed syllables - 11 in total - are included. Data on [e] and [o] have been left out in the present study.

The data in this project were drawn from recordings of three native speakers of Canton Cantonese and three of Hong Kong Cantonese. All the Hong Kong speakers were born in Hong Kong and grew up there. Two of the Canton speakers were born

Table 1. Distribution of Cantonese Vowels

ɑ:	ɑ:j	ɑ:w	ɑ:m	ɑ:n	ɑ:ŋ	ɑ:p	ɑ:t	ɑ:k
	ɒj	ɒw	ɒm	ɒn	ɒŋ	ɒp	ɒt	ɒk
ɛ:				ɛ:ŋ	ɛ:ŋ			
	ej							
					Iŋ			Ik
i:		i:w	i:m	i:n		i:p	i:t	
œ:					œ:ŋ			œ:k
	øy			øn			øt	
ɔ:	ɔ:j			ɔ:n	ɔ:ŋ		ɔ:t	ɔ:k
		ow						
					Uŋ			Uk
u:	u:j			u:n			u:t	
y:				y:n			y:t	

Table 3. Mean Formant Values in Hz. Individual Speakers for the Vowels for which a Significant Difference was found.

	i:		y:		ɛ:		ø
Speaker	F2	F3	F2	F3	F1	F2	F1
Canton 1	1935	2999	1858	2223	469	1658	391
Canton 2	2329	3085	1830	2348	496	2010	481
Canton 3	2083	3086	1871	2297	522	1783	478
Hong Kong 1	2291	3199	2059	2663	696	1957	507
Hong Kong 2	2325	3104	1975	2370	476	2035	468
Hong Kong 3	2597	3396	2270	2874	533	2169	507

Table 2. Word List

Vowel	Morpheme	English Gloss
[i:]	[ji:^{44}]	"meaning"
	[hi:m^{44}]	"to owe"
	[hi:p^{44}]	"to assist"
[I]	[p^{h}Iŋ^{44}]	"to employ"
	[pIk^{5}]	"to compel"
[ɛ:]	[mɛ:^{55}]	"to shoulder"
	[pɛ:ŋ^{44}]	"handle"
	[hɛ:k^{4}]	"to eat"
[y:]	[jy:^{33}]	"prior to"
	[hy:n^{44}]	"to advise"
	[hy:t^{4}]	"blood"
[ø]	[søn^{44}]	"letter"
	[søt^{5}]	"grasshopper"
[œ:]	[hœ:^{55}]	"boots"
	[hœ:ŋ^{44}]	"to"
	[kœ:k^{4}]	"foot"
[ɑ:]	[pɑ:^{44}]	"bully"
	[pɑ:t^{4}]	"eight"
	[pɑ:k^{4}]	"hundred"
[ɒ]	[pɒn^{44}]	"funeral"
	[pɒt^{5}]	"pen"
	[pɒk^{5}]	"north"
[ɔ:]	[pɔ:^{44}]	"spread"
	[pɔ:k^{4}]	"wide"
[U]	[p^{h}U ^{44}]	"meet"
	[pUk^{5}]	"forecast"
[u:]	[fu:^{44}]	"trousers"
	[fu:t^{4}]	"wide-"

and raised in the city of Canton, while the remaining speaker was born in Zhongshan county but had lived in Canton since three years of age. All speakers are male, their ages ranging from the mid twenties to the mid thirties.

The word list for the vowel analysis (see Table 2) consists of 28 monosyllabic morphemes, with two to three morphemes for each of the 11 vowels[1]. These morphemes were randomized in the word list, each appearing three times, yielding six to nine tokens for each vowel per speaker. Each token appears in the sentence frame: [ŋɔ: jiu tUk -- pei nei tʰɛ:ŋ] "I want read -- for your hear". All the morphemes bear level tones, and to minimize the effect of initial consonants on vowels, the initials in the syllables chosen are generally either labials or glottal consonants. There are, however, two syllables that begin with s- (which is inevitable given the phonotactic constraints on [o]), and one syllable with k-.

For each token, the frequencies of the first two formants were measured based on wide-band spectrograms. The following principles were followed in the formant measurements: whenever a clear steady state portion was discernable, the mid-point of the steady state was chosen as the point at which to measure the formants. Where no steady state portion can be observed, if the initial was h-, the point immediately before the deflection of the formants was measured. In similar circumstances, if the initial was not h-, the midpoint of the vowel was chosen instead. The frequency of F3 was also measured for the high front vowels [i:] and [y:]. Naturally, difficulties in sorting out the first two formants were encountered in respect to the back vowels (cf. Ladefoged 1967), chiefly [ɑ:], [ɔ:] and [u:], and occasionally [U] and [ø]. The vowels that presented difficulties in measurement varied from one speaker to another. In problematic cases, narrow-band sections were taken at the point where the measurements were made, to help decide on the locations of F1 and F2.

With regard to vowel duration, values were obtained for every token except those that have palatal glide initials.

Results

For each speaker, the vowels were plotted in an acoustic space with F1 along the y axis and the differences between F2 and F1 along the x axis. Distances along the axes were made proportional to the mel scale in order to approximate better the perceived distances in phonetic space. Using the ellipse program of Davis (1977), for each cluster of vowel points, an ellipse with a radius of two standard deviations were drawn along axes oriented toward the principal components. In addition to plotting the vowel chart of each speaker, a vowel chart was plotted of the pooled data of each of the two groups of speakers. These charts are given in Figures 1 through 8.

The first observation that can be made is that if the ellipses of the vowels of Canton speakers (see Figure 7) are superimposed on those of the vowels of Hong Kong speakers (see Figure 8), one would find that aside from the vowels [i:], [y:], [ɛ:], [ø], all the other vowels in one system more or less overlap with corresponding ones in the other system. This means that if systematic differences can be found in these four vowels, these differences could not be due to a consistent bias in the physiological makeup of the particular subjects chosen for the recordings. If this were the case, one would expect this bias to manifest itself in all the vowels, which is not the case with our data.

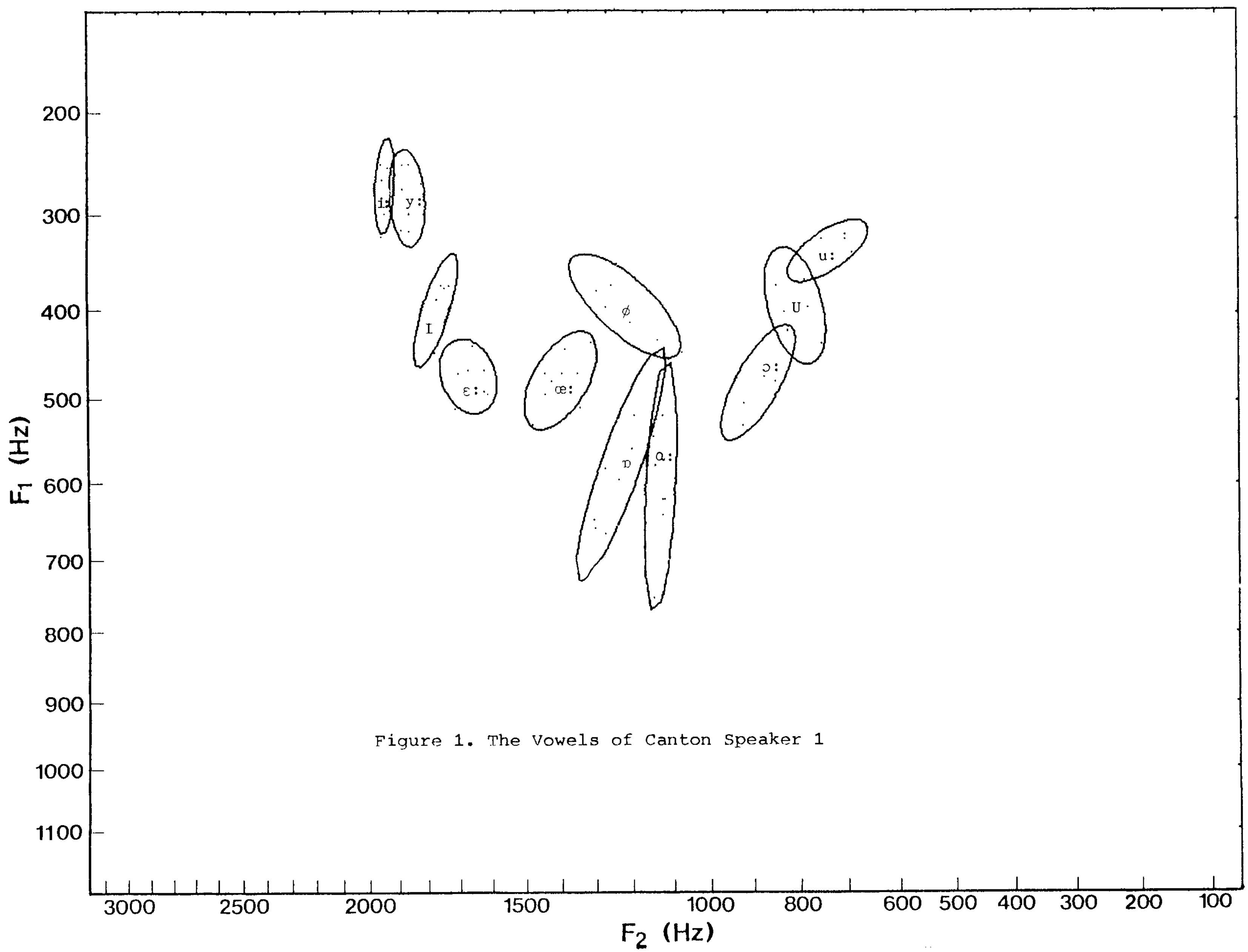


Figure 1. The Vowels of Canton Speaker 1

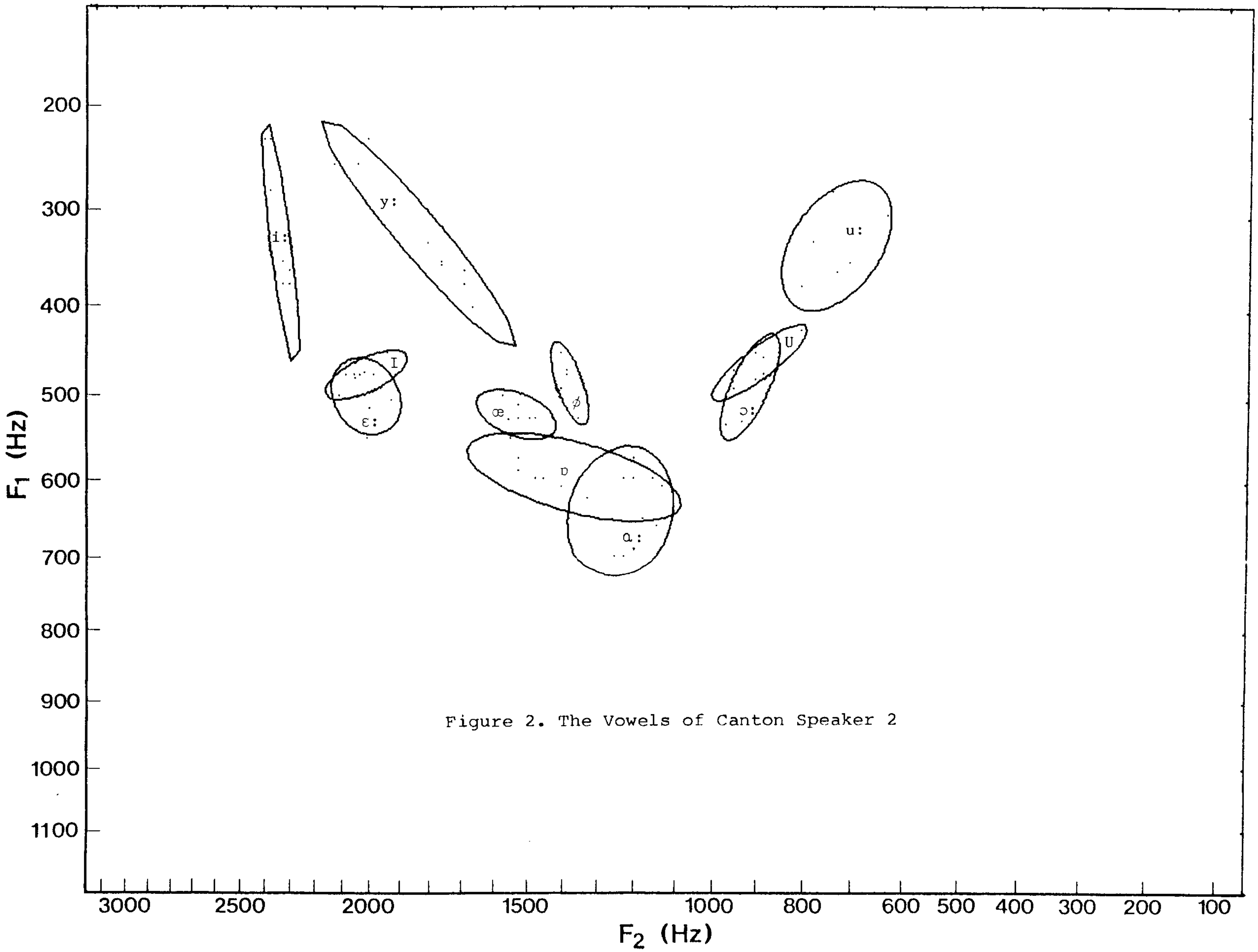


Figure 2. The Vowels of Canton Speaker 2

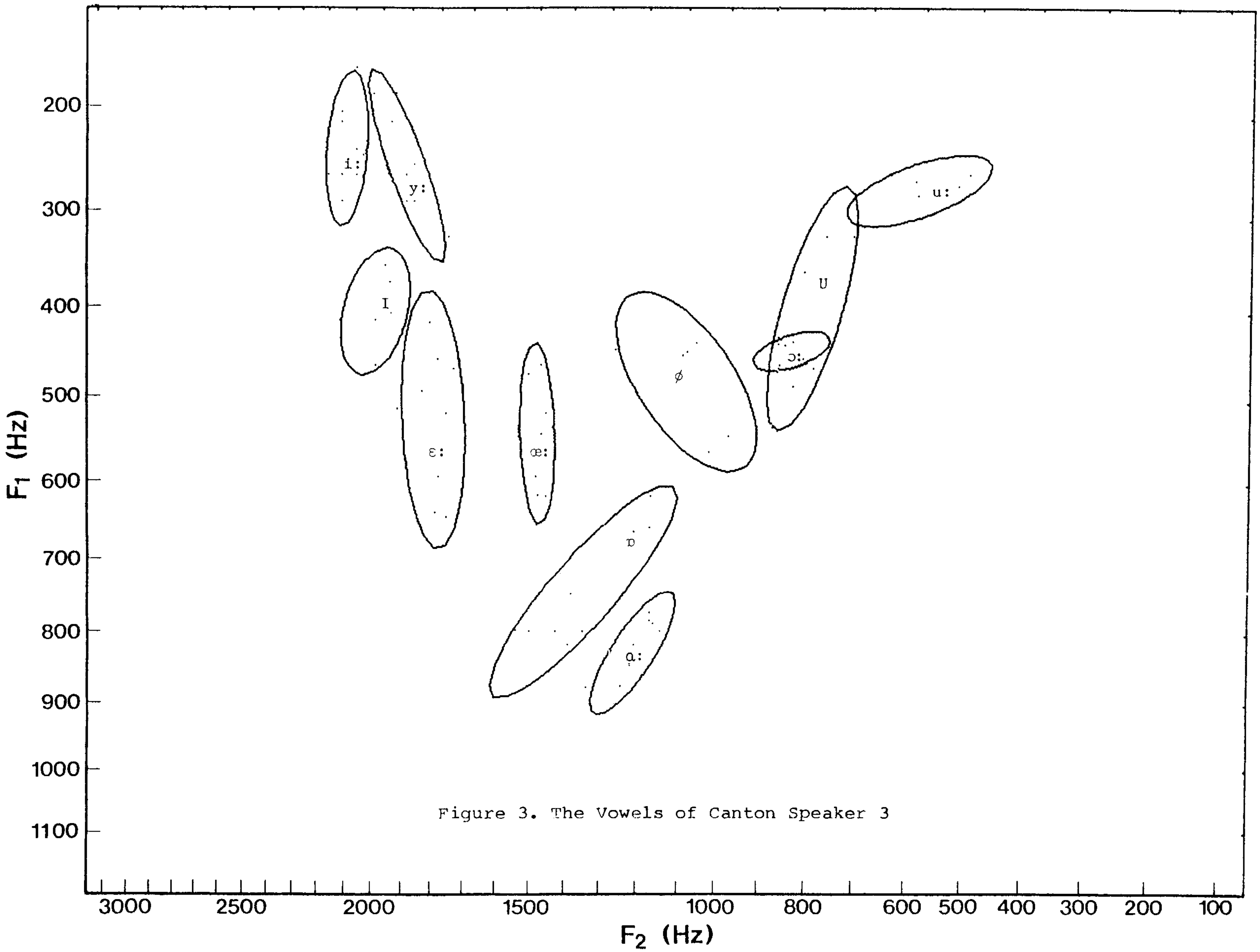


Figure 3. The Vowels of Canton Speaker 3

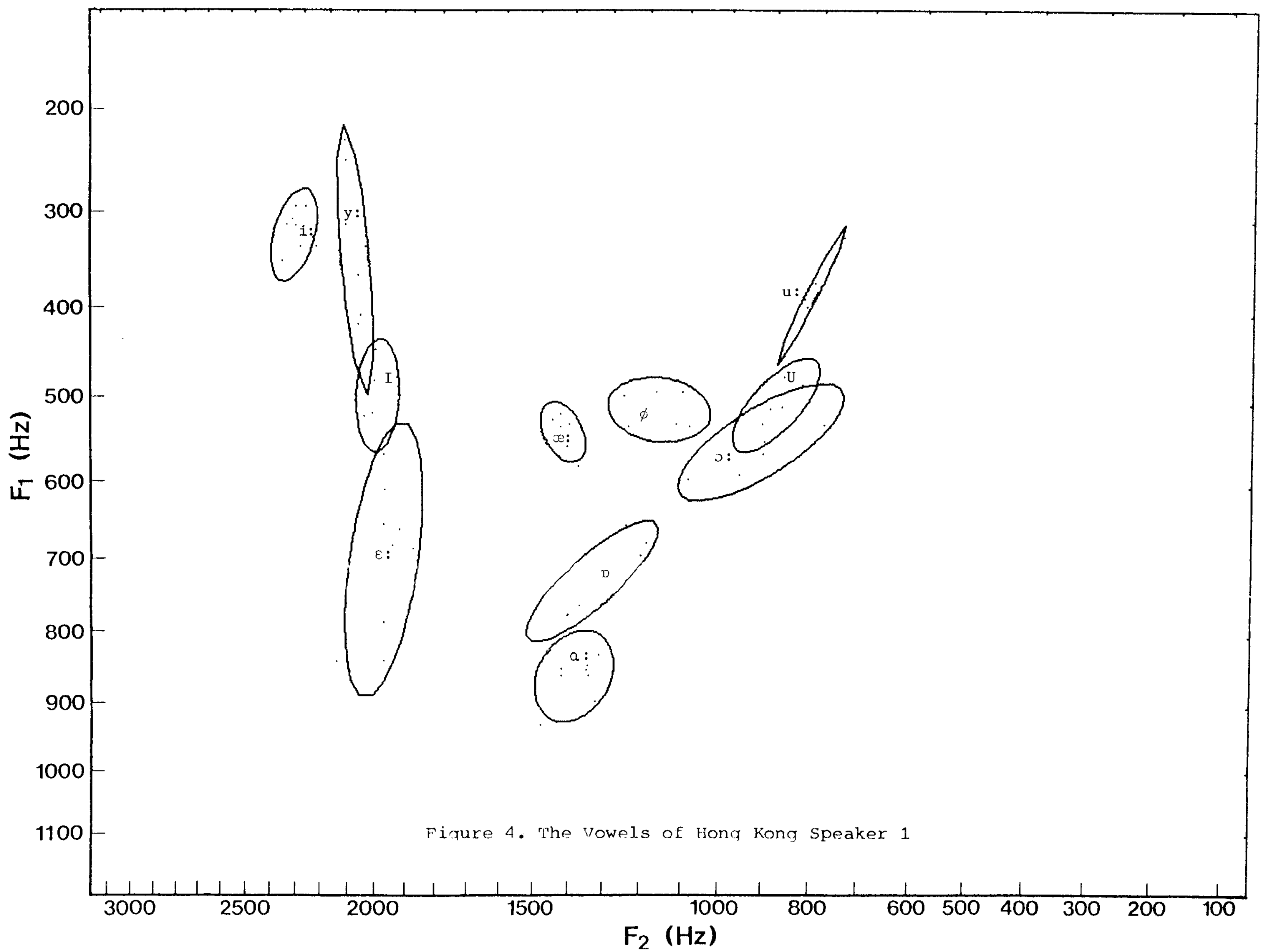


Figure 4. The Vowels of Hong Kong Speaker 1

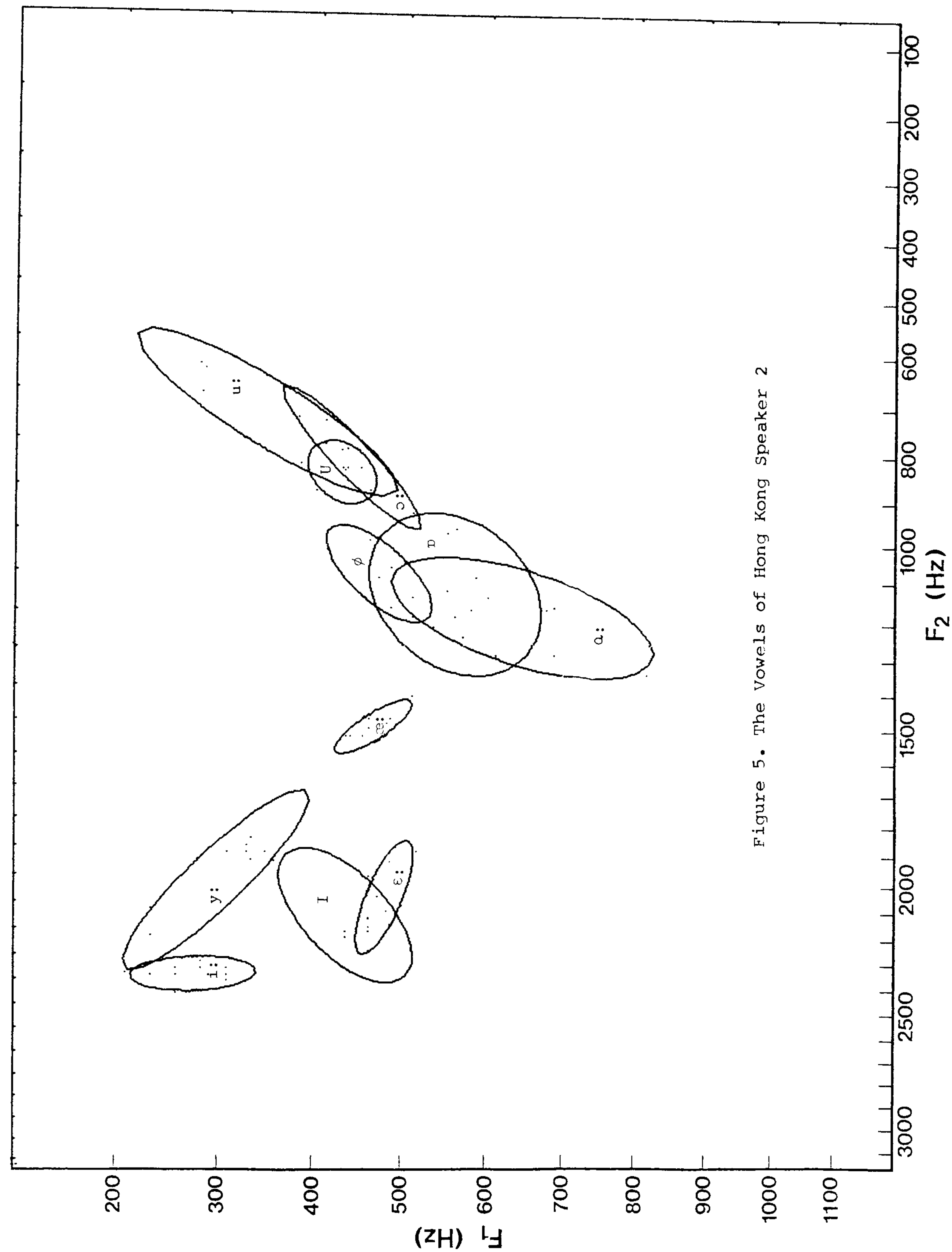


Figure 5. The Vowels of Hong Kong Speaker 2

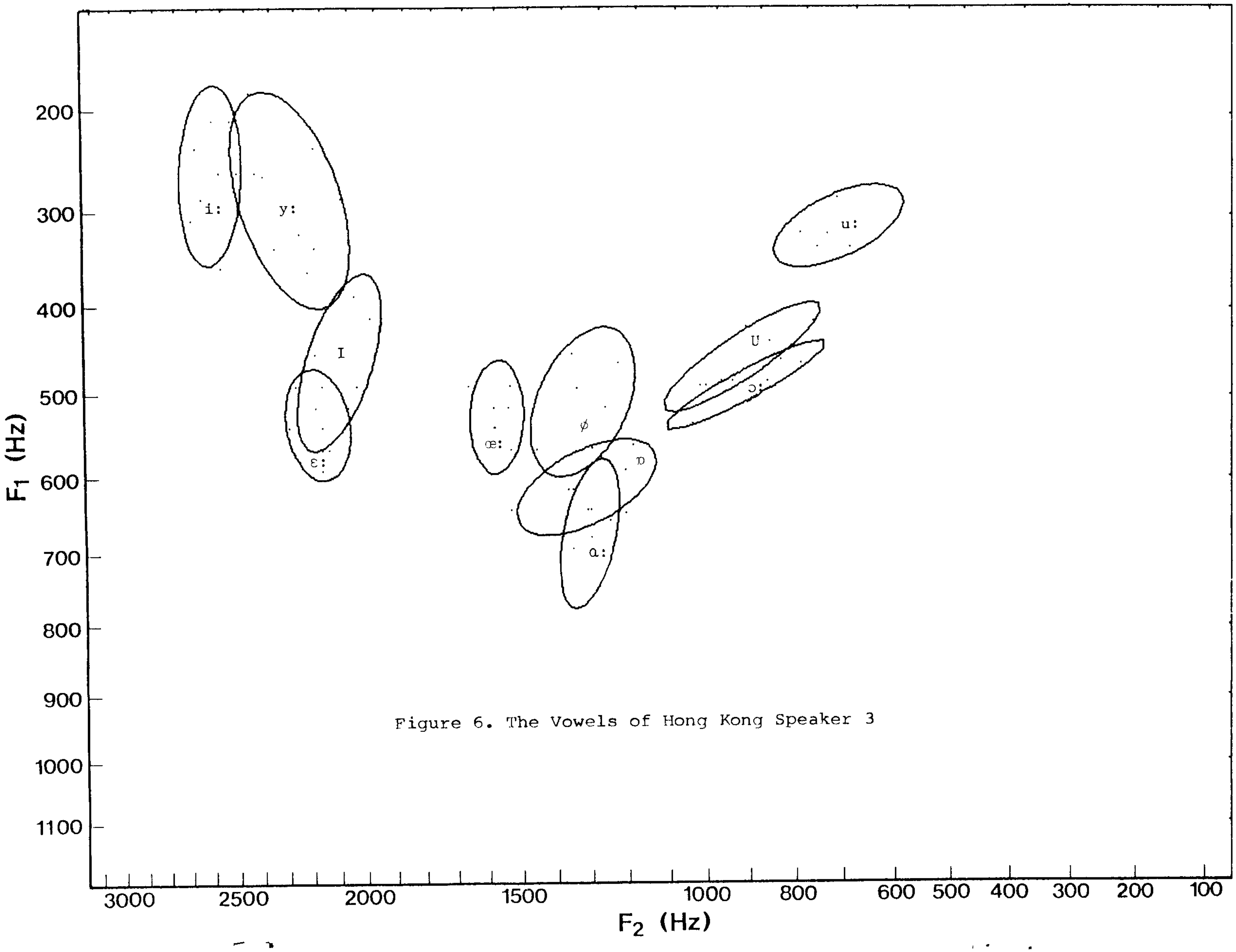


Figure 6. The Vowels of Hong Kong Speaker 3

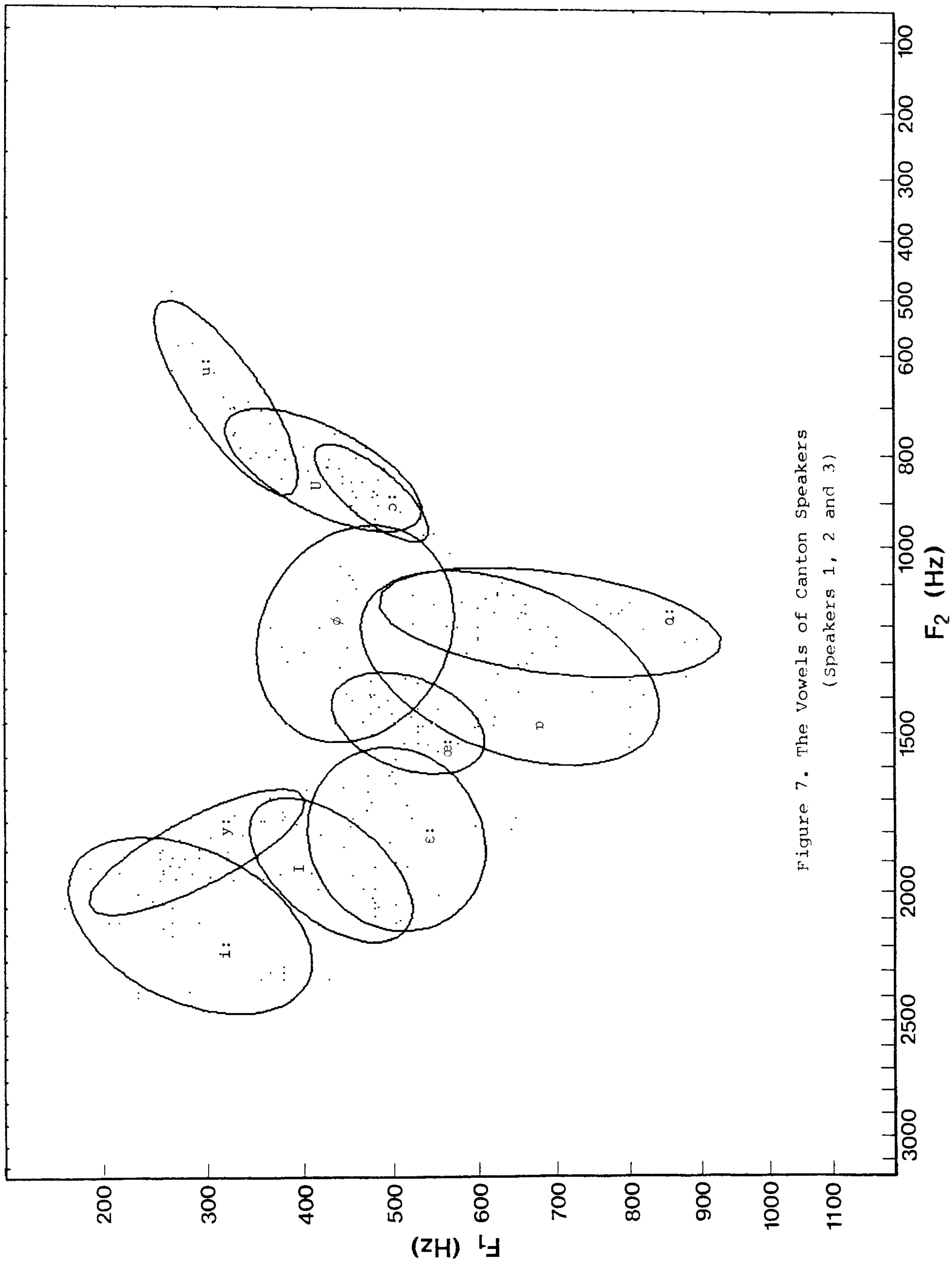


Figure 7. The Vowels of Canton Speakers
(Speakers 1, 2 and 3)

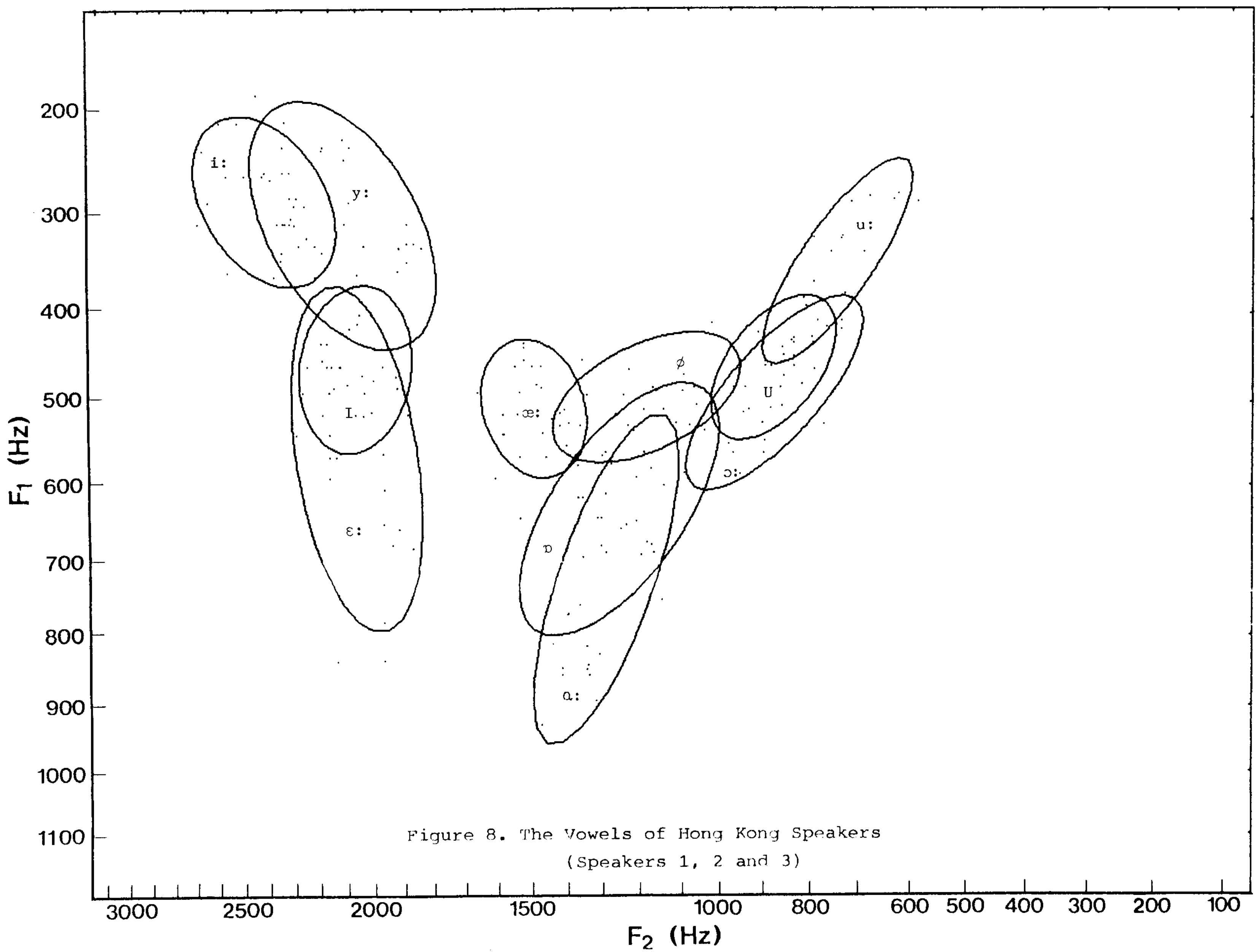


Figure 8. The Vowels of Hong Kong Speakers
(Speakers 1, 2 and 3)

With regard to the above-mentioned four vowels, the following comments can be made:

(1) The front long vowels [i:], [y:], [ɛ:] are more peripheral in Hong Kong speakers than in Canton speakers. In the case of [ɛ:], the vowel is also lower in the Hong Kong speakers. These observations of the vowels charts are supported by statistical analyses of the formant values. A grouped t-test was carried out on the differences between the two sets of speakers with respect to the F1, F2 and F3 of [i:] and [y:]. As expected, the differences in F1 for either vowel were not statistically sgnificant. However, the difference in F2 and F3 between the two groups, for both [i:] and [y:] , was highly significant ($p<.001$). The mean F2 and F3 of [i:] for Canton speakers were 2116 Hz and 3056 Hz respectively, while the corresponding values for Hong Kong speakers were 2404 Hz and 3233 Hz. The F2 and F3 of [y:] show respective mean values of 1853 Hz and 2289 Hz for Canton speakers, and 2101 Hz and 2635 Hz for Hong Kong speakers. As for the vowel [ɛ:], significant differences exist between the two groups with regard to F1 ($p<.005$) as well as F2 ($p<.001$). The mean F1 and F2 of the Hong Kong speakers were 568 Hz and 2053 Hz respectively, again higher than the corresponding figures for the Canton speakers: 496 Hz and 1817 Hz.

To see whether this difference in vowel quality between the two groups is uniform across all speakers, the mean F1 and F2 of each speaker was calculated for each vowel, as shown in Table 3. The mean values show a consistent picture for [y:] in that the mean F2 and F3 values of any of the Hong Kong speakers is greater than the corresponding value for any of the Canton speakers.For [i:], the Hong Kong speakers all have a higher mean F3 than the Canton speakers. However, Canton speaker 2 has a higher value of F2 than two of the Hong Kong speakers. As for the vowel [ɛ:], the mean F2 of Hong Kong speakers 2 and 3 is higher than all three of the Canton speakers, but Hong Kong speaker 1 has greater mean F2 over only two of the Canton speakers. In the case of F1, there is some degree of irregularity in that while two of the Hong Kong speakers (1 and 3) have higher mean F1 over any of the Canton speaers, Hong Kong speaker 2 has a lower mean F1 than two of the Canton speakers. The overall picture, then, suggests a systematic difference between the two varieties of Cantonese, bringing out the fact that Hong Kong speakers have relatively more peripheral front long vowels than their Canton counterparts.

(2) It also emerges clearly from the vowel charts that whereas the [ø] in Canton speakers differs from [oe:] in both backness and height, the former being higher and backer than the latter, the [ø] in Hong Kong speakers differs from [oe:] only in degree of backness. This observation is again buttressed by statistical analyses. A grouped t-test performed on the F1 and F2 values of [ø] reveals no significant difference between the two groups of speakers vis-a-vis F2. On the other hand, a significant difference obtains with respect to F1 ($p<.01$). The mean F1 is 450 Hz for Canton speakers, compared to the mean F1 of 494 Hz of the Hong Kong speakers. The relatively lower mean F1 value of the Canton speakers corresponds to a perceptually higher vowel height of [ø] in the Canton variety.

(3) A third consistent pattern of difference between the two vowel systems is concerned with the relative position of [ɒ] in relation to [ɑ:]. In all three Canton speakers, the [ɒ] is more front and close than [ɑ:]. On the other hand, in each of the Hong Kong speakers, the [ɒ] has the same degree of frontness as [ɑ:] and differs from the latter only in being higher in height (see Figures 1-6).

Table 4. Mean Vowel Durations of Canton and Hong Kong Speakers for Different Vowels (msec)

	i:	I	y:	ɛ:	ø	œ:	ɒ	ɑ:	ɔ:	U	u:
Canton Speakers	140	98	160	223	98	226	102	233	255	90	207
Hong Kong Speakers	123	85	158	211	81	210	85	215	241	75	199

Table 5. Range of Vowel Durations in Different Syllable Types for Different Speakers (msec)

		Canton 1	Canton 2	Canton 3	HK1	HK2	HK3
Long Vowels in Open Syllables	(Range)	249–286	189–260	320–390	320–380	240–308	220–260
	(Mean)	265	224	352	345	270	233
Long Vowels in Closed Syllables (__Nasals)	(Range)	156–220	122–194	190–326	170–296	120–198	135–198
	(Mean)	185	164	250	238	171	170
Long Vowels in Closed Syllables (__Stops)	(Range)	129–222	100–144	141–272	121–222	78–177	115–175
	(Mean)	172	127	215	162	129	148
Short Vowels in Closed Syllables (__Nasals)	(Range)	93–119	65–97	88–152	96–109	68–102	79–94
	(Mean)	100	87	126	104	86	88
Short Vowels in Closed Syllables (__Stops)	(Range)	84–111	76–82	87–108	68–94	57–76	53–109
	(Mean)	96	78	101	77	65	82

(4) A fourth systematic difference between the two varieties of the same dialect lies in vowel duration. The mean values of the durations of all the vowels for each group were computed and are given in Table 4. As the table shows, the mean vowel duration for each of the vowels is consistently higher in Canton speakers. However, a grouped t-test carried out on the difference between the two groups in the durations of each vowel shows no significant inter-group difference for most of the vowels. Significant differences were found only in the case of three short non-front vowels, [ɒ], [o] and [U]. In all these three vowels, Canton speakers show relatively longer durations ($p<.01$), as given in Table 4.

To sum up the major acoustic segmental differences between Canton Cantonese and Hong Kong Cantonese, the latter has relatively more peripheral front long vowels and shorter non-front short vowels. Systematic differences also obtain in the positions of [ø] and [ɒ] in respective relations to [oe:] and [ɑ:] in acoustic vowel space.

Implications of Present Findings for Phonemic Analysis

The acoustic findings in this study have several implications for the analysis of vowel phonemes in Cantonese. Firstly, our findings have consequences regarding the use of certain phonetic symbols to represent some of the vowels. Kao (1971) and Hashimoto (1972) both use the symbol [ø] to represent the short vowel that occurs only between dental/alveolar initials and a palatal glide or dental/alveolar endings. These authors explicitly refer to this vowel as a front rounded vowel, the rounded counterpart of [e]. Our vowel charts show that this vowel in fact resembles a central vowel in the case of Canton speakers 1 and 2, and looks like a back vowel in the case of Canton speaker 3 and all of the Hong Kong speakers. In any case, unless there is phonological evidence demonstrating that [ø] behaves as a front vowel, there should not be any motivation for describing it as such on phonetic grounds. In this connection, a phonotactic constraint on Cantonese syllable structure may lend support to classifying [ø] as a front vowel. In Cantonese, labial initials cannot occur before [y:], [ø], and [oe:] in a syllable (Hashimoto 1972:140). If [ø] is regarded as a front vowel, the statement of this phonotactic condition would be very simple: labial initials cannot occur before front rounded vowels. However, this seems to be the only piece of phonological evidence in favor of aligning [ø] with the front vowels.

Here, it is instructive to review some of the descriptions of this vowel in the literature. In one of the earliest descriptions of this vowel in modern phonetic terms, by Williams (1856), this vowel in the morphemes 津, 卒 (romanized by Williams as <u>tsun</u>, <u>tsut</u>), is said to be like "the u in <u>put</u>, <u>nuisance</u>, the sound lying between the vowel sounds in <u>fun</u> and <u>fool</u>". The romanization actually reflects the alignment of the vowel with the back series. In a much later study, Chao(1947), the symbol [ɵ],the central vowel intermediate between [e] and [o] in IPA, is used to represent this vowel. In Chao's phonemic analysis of Cantonese, this vowel is grouped together with [o] to form the short vowel phoneme /ŏ/. Yuan (1960) follows essentially Chao's analysis. Thus, it appears that the authors of the more recent studies, i.e. Kao and Hashimoto, apparently out of a concern for pattern congruity and simplicity in phonological statement, have departed from the phonetic content of this central or back vowel which they have labelled as [ø].

Our findings also shed light on the relative position of [ɒ] to [ɑ:]. In Kao (1971), [ɒ] is situated above [ɑ:] and to the front of it, in accordance with the

IPA descriptions of these symbols. In Yuan (1960), however, it is claimed that the long [ɑː] is a front [aː], with its short counterpart lying higher and to the back of it. These two analyses are given schematically below:

Yuan(1960)	Kao(1971)
[ɒ]	[ɒ]
[aː]	[ɑː]

In our vowel charts, the position of long "a" seems to point to a relatively back [ɑː]. [ɒ] is either directly above [ɑː] or above and to the front of it, but never to its back. Our findings, therefore, support Kao's analysis and contradict Yuan's.

The values on vowel duration from this study confirm Kao (1971)'s finding that length is a primary distinctive feature of the Cantonese vowel system. Table 4 gives the mean durations of all long and short vowels for both Canton and Hong Kong speakers. It is clear that the mean durations for long vowels all exceed 100 msec, whereas the mean durations for short vowels all fall below 100 msec (with the exception of [ɒ] in Canton speakers).

In terms of mean vowel durations, the following hierachy holds for all six speakers of Cantonese, as can be observed from Table 5.

duration of long vowels in open syllables
>duration of long vowels in nasal-final syllables
>duration of long vowels in stop-final syllables
>duration of short vowels in nasal-final syllables
>duration of short vowels in stop-final syllables

Table 5 further shows that for four of the speakers (Canton speakers 1 and 2; Hong Kong speakers 1 and 3), the range of vowel duration for open syllables is clearly distinct from that in closed syllables; likewise, the range of durations for long vowels in closed syllables is distinct from that for short vowels in closed syllables. In the tokens of the two remaining speakers (Canton 3 and Hong Kong 2), there is only a slight degree of overlap in the range of vowel durations of these three types of syllables.

The fact that the long/short (or tense/lax) distinction shows up so clearly in the speech of the six speakers is surprising, because the tense/lax distinction in Cantonese is non-phonemic except for the low vowels [ɒ] and [ɑː]. If we refer to Table 1, we find that tense/lax members of a vowel pair are predictable from their distributions, with the exception of the two low vowels. Thus, each of the vowel pairs [ɛː] and [e], [iː] and [I], [oeː] and [ø], [ɔː] and [o], [uː] and [U] are in complementary distribution with respect to syllable endings. For these vowels, the tense/lax distinction can be handled by phonological redundancy rules, as has been done by Hashimoto (1972). In addition, in checked syllables (syllables with stop endings) bearing non-low tones, the tense/lax distinction is tonally conditioned, so that only lax vowels cooccur with the high tone, whereas tense vowels cooccur with the mid tone (Chao 1947:24).

If the tense/lax distinction is largely redundant in the phonemic system of Cantonese, why should it persist in playing a role in the system? Figures 1-6 may suggest a partial answer. These figures demonstrate that, for about three of the speakers, the ellipses of the short vowels [I], [ɒ], and [U] overlap with those of the long vowels [ɛ:], [ɑ:], and [ɔ:] respectively. Referring to Table 1 again, one would find that [I] contrasts with [ɛ:], [ɒ] with [ɑ:], and [U] with [ɔ:]. One could also see that [I] is in complementary distribution with [i:], and [U] with [u:]. In some analyses of Cantonese vowels, e.g. Wong (1941), [I] is grouped together with [i:] , and [U] with [u:], as members of the same phoneme. One might have expected those vowels that do not contrast with each other, rather than those that do contrast, to overlap in acoustic space. But the contrary is found in our data. One reasonable hypothesis to account for this, in view of Kao's findings and ours on vowel duration, is to suggest that length plays a distinctive role in these vowel pairs, and this saves these vowels from being collapsed in the hearer's perception despite similarities in formant values.

Notes

(1) One of the morphemes in the word list, represented by the character 柄, has two readings: a colloquial reading $[\text{pɛ:ŋ}]^{44}$ and a literary reading $[\text{pIŋ}]^{24}$. In our recordings, all the subjects used the colloquial reading.

Acknowledgments

This is part of an acoustical study of Cantonese jointly undertaken by Eric Zee and myself. I am grateful to Ian Maddieson for his advice at various stages of this project. I have also benefited a great deal from discussions with Ren Hong-mo and Eric Zee.

References

Davis, S.B. (1976) "Computer evaluation of laryngeal pathology based on inverse filtering of speech" SCRL Monograph no. 13. Speech Communication Research Laboratory: Santa Barbara, CA.

Chao, Y.R. (1947) Cantonese Primer. Cambridge: Harvard University Press.

Hashimoto, Ann (1972) Cantonese (The Yue Dialects 1). Cambridge: Cambridge University Press.

Kao, Diana (1971) The Structure of the Syllable in Cantonese. Mouton: The Hague.

Ladefoged, P. (1967) Three Areas of Experimental Phonetics. London: Oxford University Press.

Williams, S. (1856) Tonic Dictionary of the Chinese Language in the Canton Dialect. Canton: Office of the Chinese Repository.

Wong, S.L. (1941) A Chinese Syllabary Pronounced According to the Dialect of Canton. Hong Kong: Zhonghua Book Company.

Yuan, Jia-hua (1960) A Synopsis of Chinese Dialects (Hanyu Fangyan Gaiyao). Beijing: The Writing Reform Publishing House.

An Inverse Filtering Study of Burmese Creaky Voice.

Hector R. Javkin and Ian Maddieson.

Paper presented at the 104th Meeting of the Acoustical Society of America.
Orlando, Florida; November 1982.

Introduction

One of the properties which the sound system of a language can vary in order to signal meaningful differences is the quality of the voice. Differences between what we call "normal voice" and creaky voice have been studied largely by examining samples produced by trained speech scientists or subjects with laryngeal pathology (cf. Laver, 1980, and the review of the relevant research on laryngeal pathology by Ludlow, 1981). To the best of our knowledge, the airflow properties of creaky voice used distinctively in a natural language have not previously been examined. This paper reports on glottal airflow for one speaker of a language which uses both these voice qualities, and discusses the advantages of the technique used.

Materials

Burmese, the national language of Burma, uses the difference between creaky and "normal" voice contrastively. Burmese is a tone language, but, as the literature shows, the tonal contrasts involve more than just differences in pitch. Several distinct properties of the four main syllable types are collectively referred to as distinctions between "tones". The following descriptions of these syllable types are based on those in the standard reference grammar of Burmese by Okell (1969). Type 1 begins on a low pitch relative to the other three tones, and usually rises. It is relatively long and it has normal voice quality. We will refer to this type as rising tone. The other three types all begin on a high pitch relative to the rising tone, and have a falling pitch contour. Among these three, Type 2 in Okell's view has normal voice quality, although some authors, e.g. Thurgood (1981: 20), suggest that this type may have a breathy voice quality. The syllable is relatively long. We will refer to this type as falling tone. Type 3 has a "glottally constricted (creaky) voice-quality, i.e. a gradually applied glottal stop", and is relatively short. We will refer to this as creaky tone. Type 4 has a final glottal stop [-ʔ], and is short.

Figure 1 shows the pitch contour and duration of these four tones in the speech of our subject, who is a male in his early twenties. The contours on the left of the figure are those found after a voiced initial consonant, the ones on the left those found after a voiceless initial consonant. Each contour represents an average of several tokens.

The words used in our study to represent the four tonal categories are given in table 1. An adaptation of the conventional transliteration of the Burmese orthography, designed to show phonemic contrasts, is used. The sequences /hl/ and /hm/ represent a voiceless alveolar lateral and a voiceless bilabial nasal respectively. In some accounts /hs/ is said to be an aspirated fricative, but it does not seem to differ from /s/ in the speech of our subject. Hence "letter" and "hungry" are homophonous.

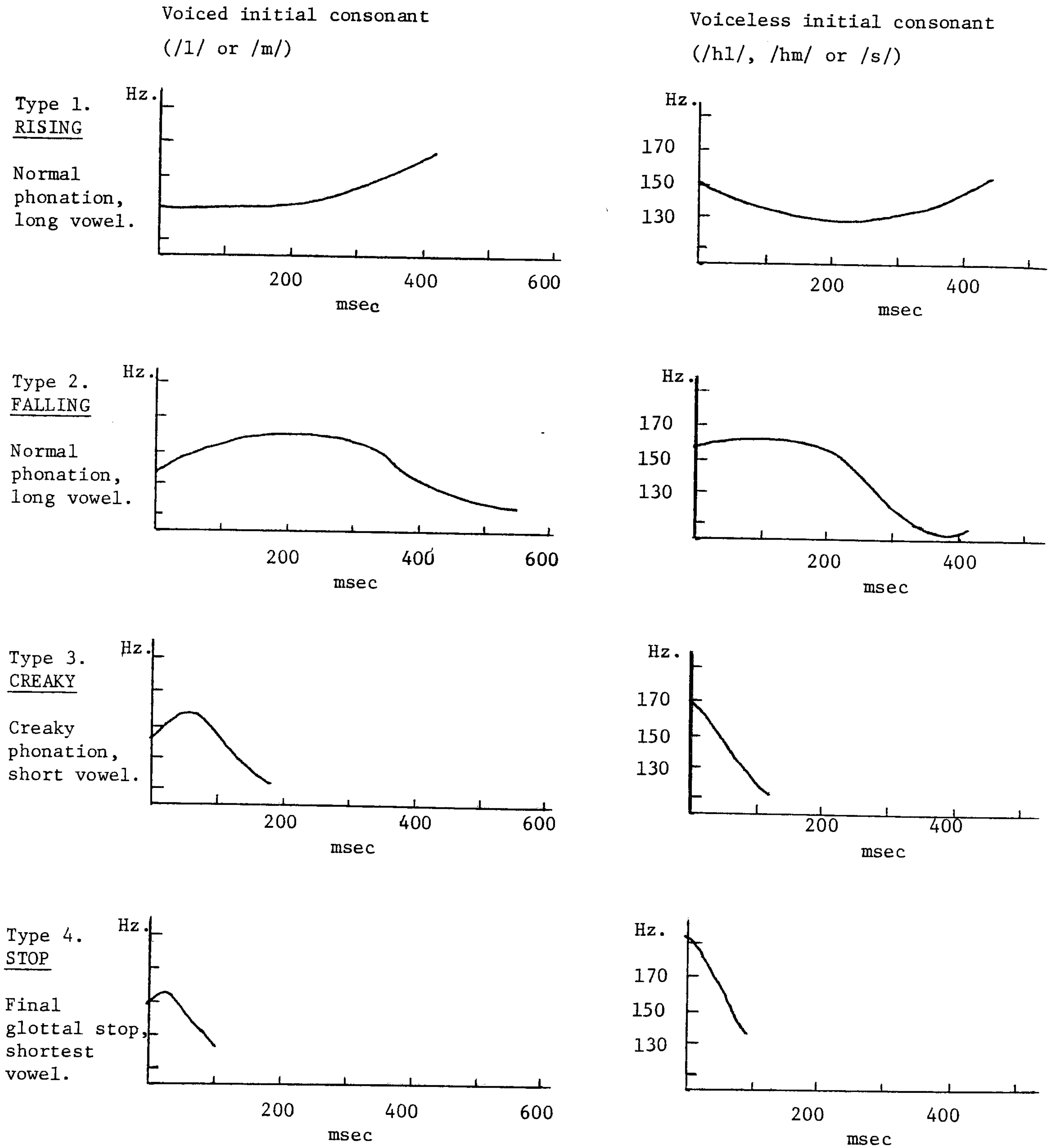


Figure 1. Pitch and duration of Burmese tones. Contour shown is the average of several tokens of representative words spoken in isolation. Contour begins at vowel onset.

Rising		Falling		Creaky		Stop	
là	"come"	lâ	"mule"	lâ	"moon"	lâʔ	"bribe"
				hlâ	"beauty"		
mà	"hard"			mâ	"lift up"		
hmà	"notice"	hmâ	"wrong"	hmâ	"from"		
sà	"letter"	sâ	"eat"	sâ	"start"	sâʔ	"pungent"
hsà	"hungry"						

Table 1. Words used in the experiment.

As will be discussed in more detail later, we are looking for characteristic differences in the glottal airflow pattern which will distinguish creaky from non-creaky phonation. However, syllables ending in a glottal stop frequently have a strong creaky component, and, as Okell remarks, the creaky tone syllables terminate in a "gradually applied glottal stop". We would therefore expect similarities between the stop tone and the creaky tone in Burmese. Thus the properties we are looking for will be ones that principally distinguish the rising and falling tones on the one hand from the creaky and stop tones on the other.

Methods.

The technique of inverse filtering was used to examine the air volume velocity in words belonging to the four tones. This technique removes the confounding influence of the vocal tract without the use of invasive procedures (Miller, 1959; Rothenberg, 1973; Fant, 1979).

A two-pole hardware inverse filter, designed by Gunnar Fant, was built at the Phonetics Laboratory at UCLA, with the variable F1 and F2 negative poles designed to cancel out resonances one-fourth the frequency of typical F1 and F2 values. This enables a recorded signal to be played through the filter at 1/4 speed.

The recording and analysis stream can be seen in Figure 2. The speaker was seated in a sound treated room and a Bruhl and Kjaer half-inch condenser microphone was held approximately 1 inch to the side of the corner of the speaker's lips while he read the 15 words shown earlier 5 times each. During the recording the ventilating system of the building was shut down in order to eliminate low frequency variations in air pressure. This precaution was necessary because of the low-frequency sensitivity of the recording and analysis procedure. The microphone output was fed through a Bruhl and Kjaer amplifier into a Tandberg FM tape recorder set at 15 inches per second. The recording was continuously monitored on an oscilloscope in order to ensure that relatively constant amplitude was maintained.

The FM tape was played back through the inverse filter at 3 and 3/4 inches per second into an Oscillomink chart recorder. Since the chart recorder has a maximum frequency response of about 1200 Hz, the fourfold reduction in playback speed increased the effective frequency response to about 4800 Hz. Figure 3 shows a sample of the three outputs which were charted: a) the speech signal, b) the speech signal with inverse filtering applied, and, c) the inverse filtered signal passed through a hardware integrator with a time constant of about 5 Hz. Because of the effects of radiation at the lips, the unintegrated inverse filtered waveform represents a differentiation of the air volume velocity. Integration

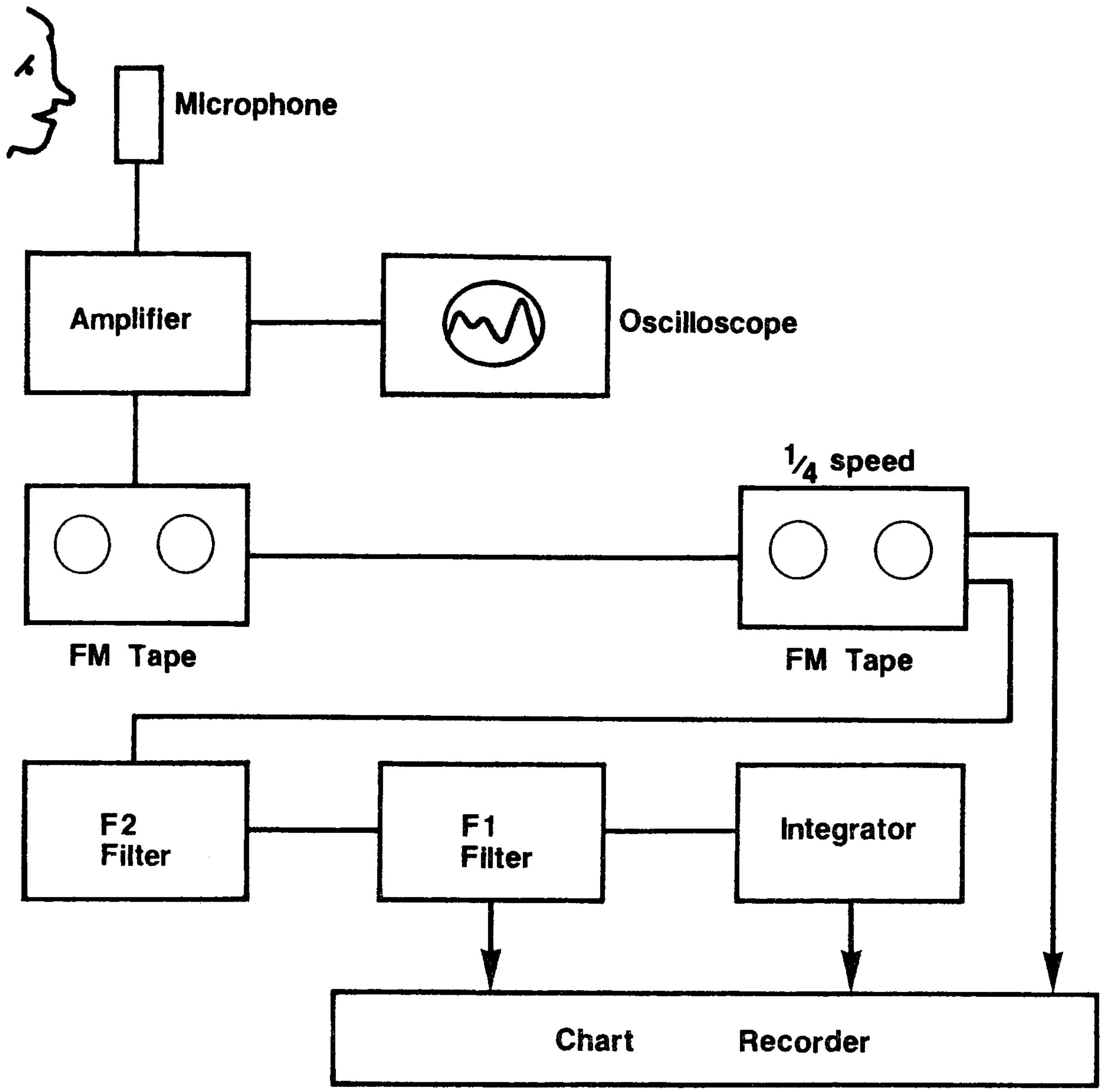


Figure 2. Diagram of recording and analysis stream for inverse filtering used at UCLA.

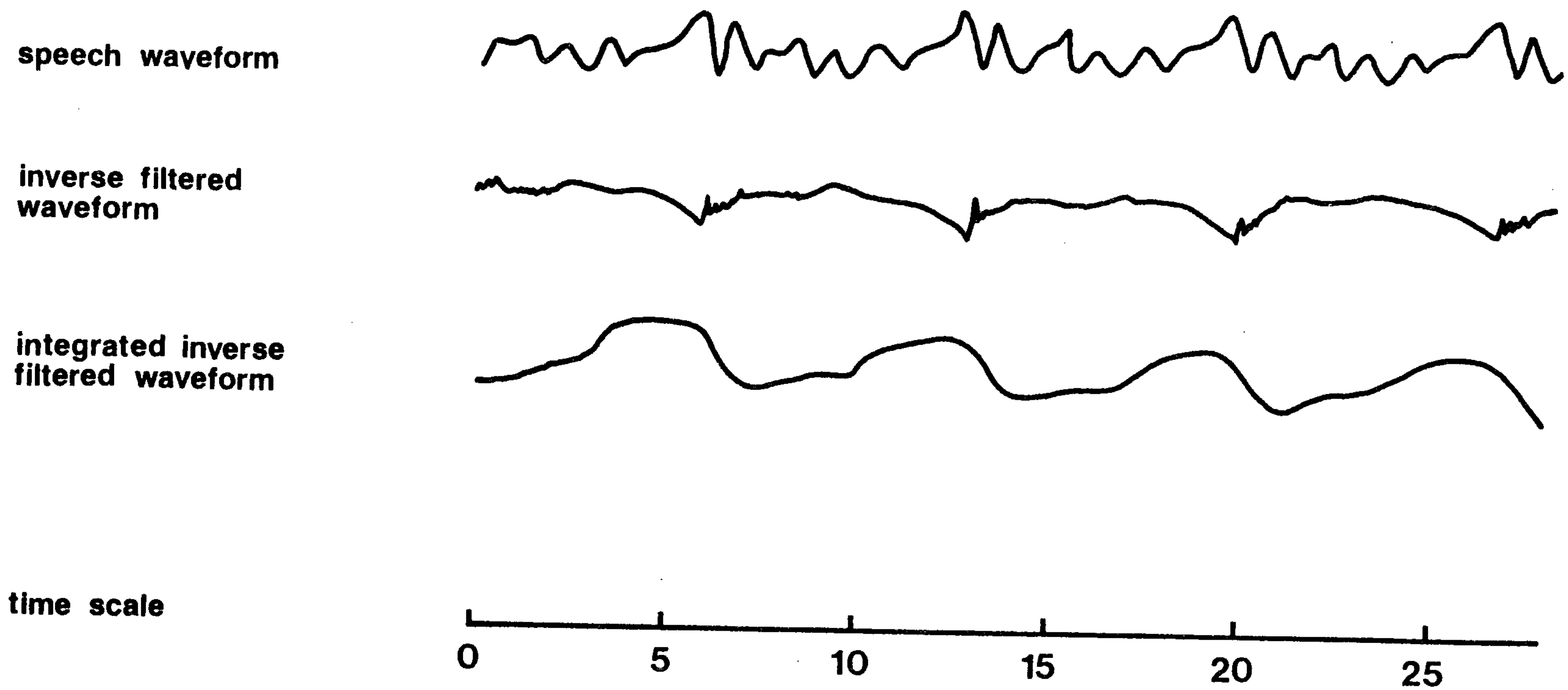


Figure 3. Sample Oscillomink output (traced directly from original and enlarged). Time scale is in msec.

provides approximately a 6 dB/octave low frequency emphasis, which offsets the 6 dB/octave high frequency emphasis caused by lip radiation. To the extent that inverse filtering is successful in removing the acoustic contribution of the vocal tract, the integrated inverse filtered signal represents the trans-glottal volume velocity.

Measurements on the Oscillomink output were made over the last 20 cycles of each token in the case of the words with rising or falling tone. Tokens of the creaky and stop words were usually too short to contain 20 cycles, so measurements were made on the highest available multiple of 5 periods. Three tokens were discarded because the signal was subject to clipping at some point in the recording and analysis chain, hence measurements were made on 72 tokens ((5 x 15) - 3).

The measurement points used for each cycle are shown in Figure 4. Point A is the beginning of the rising branch. Point B is the highest point of the cycle. Point C is the end of the falling branch. Point D is where the perpendicular from the highest point intersects a line between A and C. The distance A-C reflects the duration of the rising branch. The distance D-C reflects the duration of the falling branch. The distance B-D reflects the maximum amplitude of the cycle. The distance C-A′ gives the duration of the closed portion of the cycle, while the distance A-A′ gives the total duration of the cycle.

Approximations of the slopes of the rising and falling branches for each cycle were obtained by dividing the amplitude measure (B-D) by the rising and falling durations. This method provides only an approximation of the slopes, but it proved more reliable than attempts to measure the slope directly, since these were affected by minor errors in matching of the inverse filter to the speech. Several other ratios and sums of the measures were also calculated. In addition, the absolute difference between each period and the preceding period was obtained in order to represent period to period variation in pitch ("jitter").

Results

A stepwise discriminant analysis program was used to determine which of the measures successfully differentiated between the tones. The most successful measures were then entered into analyses of variance.

Several measures, including the rising and falling durations and the slopes, provided valid means of differentiating between some or all of the tones. In evaluating our results it is important to keep in mind that these measures, although they may also be capable of distinguishing voice quality, are related to differences in fundamental frequency and amplitude. We will attempt to assess these three factors in the following discussion. Even the three "tones" with falling pitch contours do not fall to the same extent or at the same rate. As remarked earlier, the best measure would be one that would tidily oppose the rising and falling tones (both with normal phonation) to the creaky tone and the stop tone (both of which have a laryngeal constriction). Such a grouping could not be attributed to similarities in pitch contour.

The means and standard deviations of the five measures of greatest interest are shown for each tone in Table 2. Means were calculated for each token first before a mean for each tone was obtained.

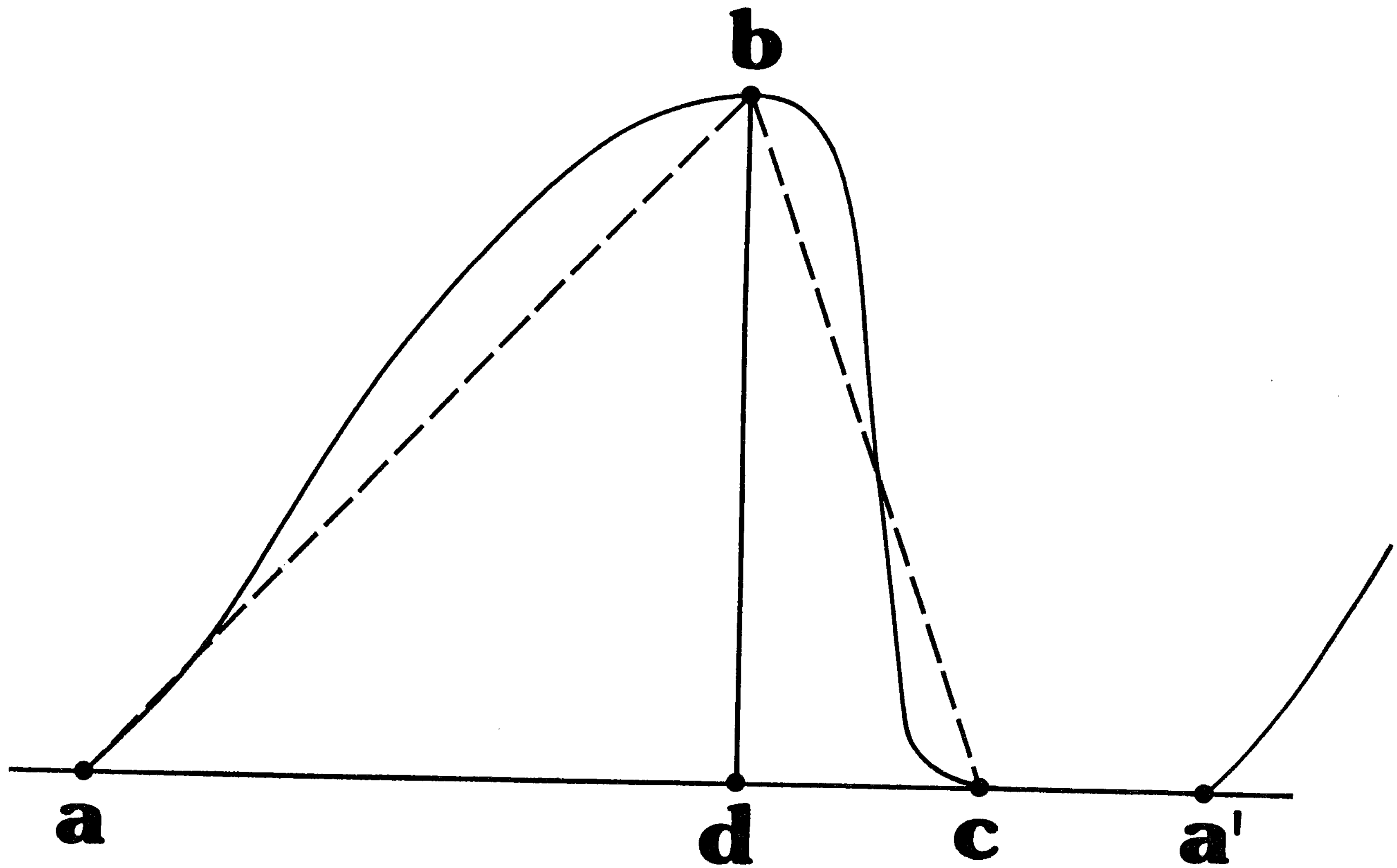


Figure 4. Measurement points on each cycle of the inverse filtered waveform. Slope of dotted lines a-b and c-b was calculated from distances a-d, d-c and b-d.

Tone	Rising branch duration $\bar{x}$	s.d.	Falling branch duration $\bar{x}$	s.d.	Rising branch slope $\bar{x}$	s.d.	Falling branch slope $\bar{x}$	s.d.	Jitter $\bar{x}$	s.d.
Rising	0.89	0.09	0.50	0.04	0.43	0.07	0.77	0.16	0.18	0.13
Falling	1.03	0.19	0.46	0.05	0.43	0.09	0.88	0.12	0.36	0.19
Creaky	0.74	0.16	0.37	0.04	0.66	0.12	1.14	0.28	0.31	0.20
Stop	0.70	0.31	0.28	0.06	0.60	0.14	1.40	0.45	0.44	0.26

Table 2. Means of five measures of glottal activity in Burmese tones. (Units are arbitrary.)

The results show the following general patterns. The duration of the rising branch of the glottal cycle tends to be shorter in the creaky and stop tones than in the rising and falling tones. The duration of the falling branch also tends to be shorter in the creaky and stop tones. As for the slopes, both the rising and falling branches of the cycle tend to be steeper in the creaky and stop tones than in the rising and falling tones. The slopes are also somewhat more variable in these tones, particularly the falling slope of the stop tone tokens. Jitter is relatively low in the rising tone, and relatively high in all three other tones. We will discuss these results in more detail in turn.

a) Rising and falling branch durations.

In thinking about the durations of the parts of the glottal cycle, the effects of the pitch contours and the word durations typical of the different tones shown in Figure 1 need to be borne in mind. The creaky and stop tones begin with high pitch and are short in duration. The periods measured on these words therefore include some relatively short ones at the beginning of the word. In the long falling tone words the periods measured do not extend to the beginning of the word. We might therefore expect the mean durations of the periods to be shorter in the shorter words because of the inclusion of the high pitched onset portion. However, the shortness of the durations of the rising and falling branches of the creaky and glottal tones cannot be simply accounted for in this way, since if this were so the durations would show not only shorter means but also greater variability. Yet the standard deviations are not larger for these tones than they are for the longer tones in which the pitch at the end of the word is more stable. In particular, this point may be made about the duration of the falling branch. This is actually longest by a small margin in the rising tone, and the standard deviation is uniformly small across all four tones. If the relatively high initial pitch level in the creaky and glottal tones was responsible for the relatively short durations of the branches of the cycle, a higher standard deviation would be expected.

It might seem paradoxical that durations of the parts of the glottal cycle show the pattern that they do. If both parts of the cycle are shorter, does it not follow that fundamental frequency is higher? In fact it is highest at the end of the rising tone, which has longer durations of the branches of the cycle than the creaky and stop tones. The paradox is resolved by considering the third component of the complete cycle, the closed portion (distance C-A′ in Figure 3).

This is usually zero in the rising tone, and hence adds no duration to the total cycle. It is much more likely to be non-zero in the creaky and stop tones, particularly in the later cycles in these tones. Thus the length of the period increases over time in these tones, but the increment occurs by adding a closed portion to the cycle.

The rising branch duration is greatest in the falling tone, the one that has been said to have a breathy quality. The difference between falling tone and the creaky and stop tones on this measure is significant. (Unless stated otherwise, significance in these results refers to a difference at at least the .01 level obtained from Tukey's studentized range test, which controls for the experimentwise error rate in the multiple comparison of means.) According to Catford (1977), in breathy voice "the vocal folds are vibrating, but without ever closing". It is possible that the slower rising pattern with the falling tone is associated with some tendency to breathiness. However, it could also be associated with low fundamental frequency in normal voice. Moreover, in the course of measurement, there were more uncertainties in determining the appropriate point to consider the beginning of the rising branch (point A on Figure 3) than were experienced with the other measurement points. We thus have less confidence in the value of the rising branch duration as a measure than we have in other measures which do not involve reference to point A.

The mean falling branch duration is significantly shorter in the stop tone than in all three other tones. It is also significantly shorter in the creaky tone than it is in the rising and falling tones. In other words, this measure both differentiates the two "glottal" tones from the tones with normal phonation, and distinguishes between the creaky and stop tones themselves. It is thus one of the most satisfactory measures for our purposes.

b) Rising and falling branch slopes.

The measures relating to the slopes of the branches also serve well to distinguish the relevant tones in an appropriate way, although in this case it is the measure associated with the rising branch that is more satisfactory than that associated with the falling branch. The slope of the rising branch is significantly steeper in the creaky and stop tones than it is in the rising and falling tones. Within each of these pairs there is no significant difference on this measure. This measure thus produces exactly the grouping of tones that was desired. However, since it is based on the rising branch duration measure, the same caveat about the reliability of the result must be entered.

The slope of the falling branch is significantly steeper for the stop tone than for the rising and falling tones, and the creaky tone is steeper than the rising tone. However, this slope is not significantly different between the creaky and falling tones, except at the much lower level of .05. Thus, as noted above, the fall duration differentiates between the relevant tones much better than the falling slope does. This result may seem paradoxical, given that the slope is calculated from the duration together with the amplitude of the pulse concerned. However, at least in the citation forms recorded in this study, there is a contour of progressively diminishing amplitude over the whole word. But, because the "tones" vary in their overall duration, this amplitude decrease is sharper in the shorter tones, contributing (among other things) to a larger standard deviation in the slope measures, especially for the falling slope. Several methods were tried to remove the effect of the amplitude contour, but because it is non-uniform across the tones, it proved difficult to do so without introducing other distortions. This measure is thus less satisfactory than the falling branch duration.

c) Jitter.

Our measure of jitter basically separates the three tones with falling pitch contours from the rising tone. Because it has the most extreme value on this measure, the stop tone is significantly different from the rising tone, but no other pairs of tones are significantly different. It is interesting to note that the falling tone has a greater amount of jitter than the creaky tone. Jitter thus fails to make the required classification of the tones.

It may be important to stress that the measures relating to jitter used in our study were not valuable in distinguishing between the phonation types. Jitter has received a lot of attention in the literature as an indication of vocal harshness (e.g. Takahashi & Koike, 1975) and has been much discussed as a likely correlate of laryngeally constricted phonation. Fourcin (1981) states that "creaky voice is characterized by its low irregular pitch" (emphasis ours). In the data we have examined, jitter is more strongly indicative of the direction (and perhaps the amount) of pitch movement, or of the absolute pitch range in use in an utterance, rather than of creaky voice. It is possible that jitter would usually differentiate phonation types if sustained phonation on a monotone is being examined (though other authors, e.g. Horyii, 1979, have challenged this), but it may be unable to do so if natural speech tokens with changing pitch are the object of investigation. Alternatively, the natural speech tokens being examined in this study may differ in this respect from the speech samples studied by others. It is not unusual for a certain amount of "creakiness" to be heard as a concomitant of low pitch. Our study shows that the perceived creakiness associated with low pitch may be typified by jitter, but that there are phonation type parameters that are independent of this characteristic.

Discussion.

Other studies of phonation, such as that of Fant (1979), indicate that the slope of the falling branch, and the associated velocity of vocal fold closure, are the most important factors in the contribution of phonation to the acoustic signal. A sharper falling slope is associated with the greater intensity of the higher part of the spectrum and greater overall acoustic energy. These are characteristic of creaky phonation. Our finding that the duration of the falling branch of the glottal cycle is one of two very satisfactory measures for discriminating between the tones which differ in phonation type in Burmese seems related to these observations. However, because natural speech tokens do not contain sustained steady-state vowels, we found that the direct measure of the falling slope was not a satisfactory discriminator. Further research is recommended to see if the slope of the rising branch proves to be a reliable indicator of phonation type elsewhere.

Conclusion

The outcomes of our study include the following: We have shown that in dealing with natural speech tokens it is important to separate the effect of pitch and intensity differences from differences in the mode of phonation. We have shown, moreover, that the reported relation between glottal stop and creaky voice can be clearly seen for our speaker; the same parameters that distinguish creaky voice and normal voice also distinguish syllables ending in glottal stop. Finally, we have shown that the inverse filtering technique can be successfully applied to the investigation of phonation types in languages where they are used contrastively. The fact that this technique is non-invasive makes it especially useful in any case where we hope to encourage our subjects to visit our laboratory again.

References

Fant, G. 1979. Vocal source analysis - a progress report. Quarterly Progress and Status Report, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm 3-4/1979: 31-54.

Fourcin, A.J. 1981. Laryngographic assessment of phonatory function. in Ludlow & Hart 1981: 116-124.

Horyii, Y. 1979. Fundamental frequency perturbation observed in sustained phonation. Journal of Speech and Hearing Research.

Laver, J. 1980. The Phonetic Description of Voice Quality (Cambridge Studies in Linguistics 31). Cambridge University Press, Cambridge.

Ludlow, C.L. 1981. Research needs for the assessment of phonatory function. in Ludlow & Hart 1981: 3-8.

Ludlow, C.L. and M. O. Hart (eds.) Proceedings of the Conference on the Assessment of Vocal Fold Pathology (Bethesda, Maryland; April 1983). The American Speech-Language-Hearing Association, Rockville, Maryland.

Miller, R.L. 1959. The nature of the vocal cord wave. Journal of the Acoustical Society of America 31: 667-677.

Okell, J. 1969. A Reference Grammar of Colloquial Burmese. Oxford University Press, London.

Rothenberg, M. 1973. A new inverse filtering technique for deriving the glottal air flow waveform during voicing. Journal of the Acoustical Society of America 53: 1632-1645.

Takahashi, H. & Y. Koike. 1975. Some perceptual dimensions and acoustical correlates of pathologic voices. Acta Oto-Laryngologica 338: 1-24.

Thurgood, G. 1981. Notes on the Origins of the Burmese Creaky Tone (Monumenta Serindica 9). Institute for the Study of the Languages and Cultures of Asia and America, Tokyo.

Index to UCLA Working Papers in Phonetics

Volumes 43-57

Anderson, Stephen R. (1980) "Why Phonology Isn't 'Natural.'" UCLA Working Papers in Phonetics 51: 36-93.

Bhaskararao, Peri (1983) "Non-Quantal Aspects of Consonant Production: A Study of Retroflex Consonants." UCLA Working Papers in Phonetics 57: 11. (with Peter Ladefoged)

Berinstein, Ava E. (1979) "A Cross-Linguistic Study on the Perception and Production of Stress." UCLA Working Papers in Phonetics 47.

Bladon, Anthony (1982) "Attempts by Human Speakers to Reproduce Fant's Nomograms." UCLA Working Papers in Phonetics 54: 40-56. (with Peter Ladefoged)

Condax, I. D. (1979) "The Epiglottis as an Articulator." UCLA Working Papers in Phonetics 45: 60-83. (with Asher Laufer)

Davey, Anthony (1982) "Liquids in Chaga." UCLA Working Papers in Phonetics 54: 93-108. (with Lioba Moshi and Ian Maddieson)

Disner, Sandra F. (1980) "Insights on Vowel Spacing: Results of a Language Survey." UCLA Working Papers in Phonetics 50: 70-92.

Disner, Sandra F. (1982) "Stress Evaluation and Voice Lie Detection: A Review." UCLA Working Papers in Phonetics 54: 78-92.

Fordyce, James F. (1980) "On the Nature of Glottalic and Laryngelaized Consonant and owel Systems." UCLA Working Papers in Phonetics 50: 120-154.

Godinez, Jr. Manuel (1978) "A Survey of Spanish and Portuguese Phonetics." UCLA Working Papers in Phonetics 44.

Greenberg, Steven (1979) "On the Perception of Contour Tones." UCLA Working Papers in Phonetics 45: 150-164. (with Eric Zee)

Greenberg, Steven (1980) "Temporal Neural Coding of Pitch and Vowel Quality." UCLA Working Papers in Phonetics 52.

Hargus, Sharon (1980) "Pitch in Sherpa." UCLA Working Papers in Phonetics 51: 132-143. (with Ian Maddieson and Jonas N.A. Nartey)

Harshman, Richard (1979) "Formant Frequencies and Movements of the Tongue." UCLA Working Papers in Phonetics 45: 39-52. (with Peter Ladefoged)

Huffman, Marie (1983) "Patterns in Allophone Distribution for Voiced and Voiceless Stops." UCLA Working Papers in Phonetics 57: 61. (with Patricia Keating and Wendy Linker)

Jacobson, Leon C. (1978) "DhoLuo Vowel Harmony: A Phonetic Investigation." UCLA Working Papers in Phonetics 43.

Javkin, Hector (1983) "An Inverse Filtering Analysis of Burmese Creaky Voice." UCLA Working Papers in Phonetics 57: 115. (with Ian Maddieson)

Keating, Patricia A. (1982) "Physiological Motivations for Phonetic Naturalness." UCLA Working Papers in Phonetics 54: 57-65. (with Wendy Linker)

Keating, Patricia A. (1983) "Phonetic and Phonological Representation of Stop Consonant Voicing." UCLA Working Papers in Phonetics 57: 26.

Keating, Patricia A. (1983) "Patterns in Allophone Distribution for Voiced and Voiceless Stops." UCLA Working Papers in Phonetics 57: 61. (with Marie Huffman and Wendy Linker)

Ladefoged, Jenny (1980) "The Ability of Listeners to Identify Voices." UCLA Working Papers in Phonetics 49: 42-50. (with Peter Ladefoged)

Ladefoged, Peter (1979) "What are Linguistic Sounds Made of?" UCLA Working Papers in Phonetics 45: 1-24.

Ladefoged, Peter (1979) "Articulatory Parameters." UCLA Working Papers in Phonetics 45: 25-31.

Ladefoged, Peter (1979) "Prediction of Vocal Tract Shapes in Utterances." UCLA Working Papers in Phonetics 45: 32-38. (with Mona Lindau)

Ladefoged, Peter (1979) "Formant Frequencies and Movements of the Tongue." UCLA Working Papers in Phonetics 45: 39-52. (with Richard Harshman)

Ladefoged, Peter (1979) "Where Does the Vocal Tract End?" UCLA Working Papers in Phonetics 45: 53-59. (with Jim Wright and Wendy Linker)

Ladefoged, Peter (1980) "The Phonetic Inadequacy of Phonological Specifications of Clicks." UCLA Working Papers in Phonetics 49: 1-27. (with Tony Traill)

Ladefoged, Peter (1980) "Instrumental Phonetic Fieldwork." UCLA Working Papers in Phonetics 49: 28-41. (with Tony Traill)

Ladefoged, Peter (1980) "The Ability of Listeners to Identify Voices." UCLA Working Papers in Phonetics 49: 42-50. (with Jenny Ladefoged)

Ladefoged, Peter (1980) "Cross-Linguistic Studies of Speech Production." UCLA Working Papers in Phonetics 51: 94-104.

Ladefoged, Peter (1982) "The Linguistic Use of Different Phonation Types." UCLA Working Papers in Phonetics 54: 28-39.

Ladefoged, Peter (1982) "Attempts by Human Speakers to Reproduce Fant's Nomograms." UCLA Working Papers in Phonetics 54: 40-56. (with Anthony Bladon)

Ladefoged, Peter (1982) "Postscript to 'Attempts by Human Speakers to Reproduce Fant's Nomograms.'" UCLA Working Papers in Phonetics 54: 109.

Ladefoged, Peter (1983) "The Limits of Biological Explanation in Phonetics." UCLA Working Papers in Phonetics 57: 1.

Ladefoged, Peter (1983) "Non-Quantal Aspects of Consonant Production: A Study of Retroflex Consonants." UCLA Working Papers in Phonetics 57: 11. (with Peri Bhaskararao)

Laufer, Asher (1979) "The Epiglottis as an Articulator." UCLA Working Papers in Phonetics 45: 60-83. (with I. D. Condax)

Lee, Thomas (1983) "An Acoustical Study of the Register Distinction in Mon." UCLA Working Papers in Phonetics 57: 79.

Lee, Thomas (1983) "The Vowel System in Two Varieties of Cantonese." UCLA Working Papers in Phonetics 57: 97.

Lindau, Mona (1979) "Prediction of Vocal Tract Shapes in Utterances." UCLA Working Papers in Phonetics 45: 32-38. (with Peter Ladefoged)

Lindau, Mona (1980) "Phonetic Differences in Nigerian Languages." UCLA Working Papers in Phonetics 51: 105-113.

Lindau, Mona (1980) "The Story of /r/." UCLA Working Papers in Phonetics 51: 114-119.

Lindau, Mona (1982) "Phonetic Differences in Glottalic Consonants." UCLA Working Papers in Phonetics 54: 66-77.

Linker, Wendy (1979) "Where Does the Vocal Tract End?" UCLA Working Papers in Phonetics 45: 53-59. (with Peter Ladefoged and Jim Wright)

Linker, Wendy (1980) "A Cross-Linguistic Study of Lip Position in Vowels." UCLA Working Papers in Phonetics 51: 1-35.

Linker, Wendy (1982) "Physiological Motivations for Phonetic Naturalness." UCLA Working Papers in Phonetics 54: 57-65. (with Patricia Keating)

Linker, Wendy (1982) "Articulatory and Acoustic Correlates of Labial Activity in Vowels: A Cross-Linguistic Study." UCLA Working Papers in Phonetics 56.

Linker, Wendy (1983) "Patterns in Allophone Distribution for Voiced and Voiceless Stops." UCLA Working Papers in Phonetics 57: 61. (with Marie Huffman and Patricia Keating)

Maddieson, Ian (1979) "Tone Spacing: Evidence from Bilingual Speakers." UCLA Working Papers in Phonetics 45: 84-88.

Maddieson, Ian (1979) "More on the Representation of Contour Tones." UCLA Working Papers in Phonetics 45: 89-92.

Maddieson, Ian (1979) "Tones and Tone Sandhi in Shanghai: Phonetic Evidence and Phonological Analysis." UCLA Working Papers in Phonetics 45: 93-129. (with Eric Zee)

Maddieson, Ian (1980) "UPSID: The UCLA Phonological Segment Inventory Database." UCLA Working Papers in Phonetics 50: 4-56.

Maddieson, Ian (1980) "Phonological Generalizations from the UCLA Phonological Segment Inventory Database." UCLA Working Papers in Phonetics 50: 57-68.

Maddieson, Ian (1980) "A Survey of Liquids." UCLA Working Papers in Phonetics 50: 93-112.

Maddieson, Ian (1980) "Vocoid Approximants in the World's Languages." UCLA Working Papers in Phonetics 50: 113-119.

Maddieson, Ian (1980) "Palato-Alveolar Affricates in Several Languages." UCLA Working Papers in Phonetics 51: 120-126.

Maddieson, Ian (1980) "Dimensions of Tone Systems." UCLA Working Papers in Phonetics 51: 127-131.

Maddieson, Ian (1980) "Pitch in Sherpa." UCLA Working Papers in Phonetics 51: 132-143. (with Sharon Hargus and Jonas N.A. Nartey)

Maddieson, Ian (1981) "UPSID: UCLA Phonological Segment Inventory Database." UCLA Working Papers in Phonetics 53.

Maddieson, Ian (1982) "Liquids in Chaga." UCLA Working Papers in Phonetics 54: 93-108. (with Anthony Davey and Lioba Moshi)

Maddieson, Ian (1983) "An Inverse Filtering Analysis of Burmese Creaky Voice." UCLA Working Papers in Phonetics 57: 115. (with Hector Javkin)

Moshi, Lioba (1982) "Liquids in Chaga." UCLA Working Papers in Phonetics 54: 93-108. (with Anthony Davey and Ian Maddieson)

Nartey, Jonas N.A. (1979) "A Study in Phonemic Universals - Especially Concerning Fricatives and Stops." UCLA Working Papers in Phonetics 46.

Nartey, Jonas N.A. (1980) "Bibliography of X-Ray Studies of Speech." UCLA Working Papers in Phonetics 49: 51-69.

Nartey, Jonas N.A. (1980) "Abstract of A Study in Phonemic Universals, Especially Concerning Fricatives and Stops." UCLA Working Papers in Phonetics 50: 69.

Nartey, Jonas N.A. (1980) "Pitch in Sherpa." UCLA Working Papers in Phonetics 51: 132-143. (with Ian Maddieson and Sharon Hargus)

Nartey, Jonas N.A. (1982) "On Fricative Phones and Phonemes: Measuring the Phonetic Differences Within and Between Languages." UCLA Working Papers in Phonetics 55.

Papcun, George (1980) "How Do Different Speakers Say the Same Vowels?" UCLA Working Papers in Phonetics 48.

Traill, Tony (1980) "The Phonetic Inadequacy of Phonological Specifications of Clicks." UCLA Working Papers in Phonetics 49: 1-27. (with Peter Ladefoged)

Traill, Tony (1980) "Instrumental Phonetic Fieldwork." UCLA Working Papers in Phonetics 49: 28-41. (with Peter Ladefoged)

Van Heuven, Vincent J. (1980) "The Phonetic Function of Rise and Decay Time in Speech Sounds: A Preliminary Investigation." UCLA Working Papers in Phonetics 49: 70-78.

Wingate, Anne (1982) "A Phonetic Answer to a Phonological Question." UCLA Working Papers in Phonetics 54: 1-27.

Wright, Jim (1979) "Where Does the Vocal Tract End?" UCLA Working Papers in Phonetics 45: 53-59. (with Peter Ladefoged and Wendy Linker)

Zee, Eric (1979) "Tones and Tone Sandhi in Shanghai: Phonetic Evidence and Phonological Analysis." UCLA Working Papers in Phonetics 45: 93-129. (with Ian Maddieson)

Zee, Eric (1979) "Effect of Vowel Quality on Perception of Nasals in Noise." UCLA Working Papers in Phonetics 45: 130-149.

Zee, Eric (1979) "On the Perception of Contour Tones." UCLA Working Papers in Phonetics 45: 150-164. (with Steven Greenberg)

Zee, Eric (1980) "Peak Intraoral Air Pressure in [p] as a Function of F_o." UCLA Working Papers in Phonetics 49: 79-89.

Zee, Eric (1980) "The Effect of Aspiration on the F_o of the Following Vowel in Cantonese." UCLA Working Papers in Phonetics 49: 90-97.

Zee, Eric (1980) "A Spectrographic Investigation of Mandarin Tone Sandhi." UCLA Working Papers in Phonetics 49: 98-116.