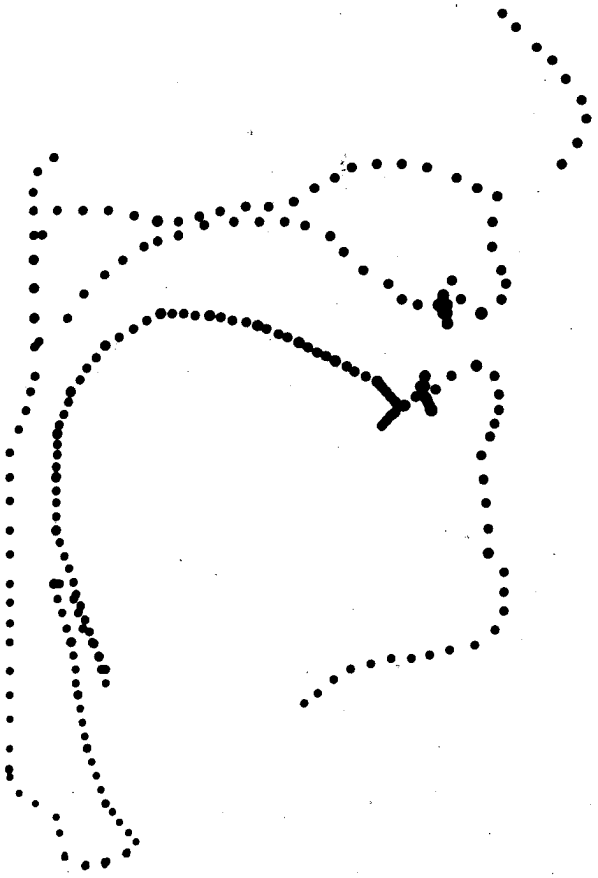
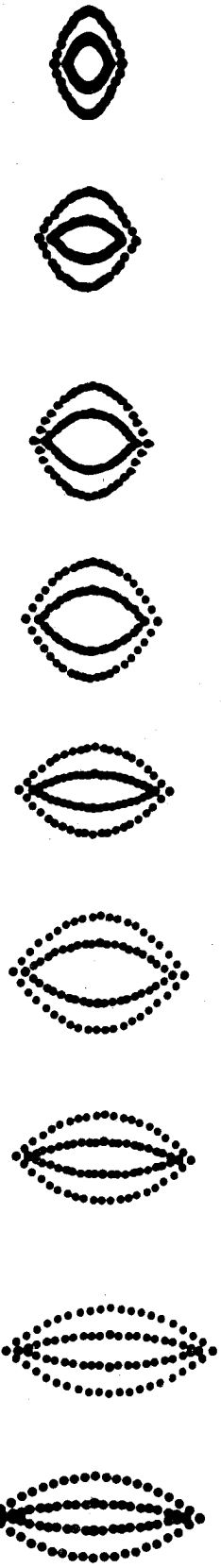




UCLA

WPP 45

March 1979

ok by BB

UCLA Working Papers in Phonetics 45

March 1979

Peter Ladefoged	What are linguistic sounds made of?	1
Peter Ladefoged	Articulatory parameters	25
Peter Ladefoged Mona Lindau	Prediction of vocal tract shapes in utterances	32
Peter Ladefoged Richard Harshman	Formant frequencies and movements of the tongue	39
Peter Ladefoged Jim Wright Wendy Linker	Where does the vocal track end?	53
Asher Laufer I. D. Condax	The epiglottis as an articulator	60
Ian Maddieson	Tone spacing: evidence from bi-lingual speakers	84
Ian Maddieson	More on contour tone features	89
Eric Zee Ian Maddieson	Tones and tone sandhi in Shanghai: phonetic evidence and phonological analysis	93
Eric Zee	Effect of vowel quality on perception of nasals in noise	130
Steven Greenberg Eric Zee	On the perception of contour tones	150
Jonas N. A. Nartey	Index of publications by members of the UCLA Phonetics Laboratory October 1975 - September 1978	165

The UCLA Phonetics Laboratory Group

Ron Carlson
Pat Coady
Sandy Ferrari Disner
Vivian Flores
Vicki Fromkin
Manuel Godinez
Steven Greenberg
Richard Janda
Hector Javkin
Peter Ladefoged
Mona Lindau-Webb
Wendy Linker
Ian Maddieson
Willie Martin
Anna Meyer
Jonas Nartey
George Papçun
Lloyd Rice
Diane Ridley
Vincent van Heuven
Renee Wellin
Anne Wingate
Jim Wright
Eric Zee

As on previous occasions, the material which is presented here is simply a record for our own use, a report as required by the funding agencies, and a preliminary account of work in progress.

Funds for the UCLA Phonetics Laboratory are provided through:

USPHS grant NS 9780
NSF grant BNS78-07680
and the UCLA Department of Linguistics

Correspondence concerning this series should be addressed to:

Phonetics Laboratory
Department of Linguistics
UCLA
Los Angeles, California 90024

What are linguistic sounds made of?

Peter Ladefoged

[Presidental address to the Linguistic Society of America,

December 1978]

When we give a description of a spoken language, what are the linguistic phonetic parameters? I want to suggest that these are not things like features. Instead I will establish that they are things like formant frequencies or parameterized vocal tract shapes. Despite the conventional wisdom, we cannot be content with specifications of linguistic phenomena in terms of physical scales representing features such as those proposed by Chomsky and Halle (1968). Even the considerably better phonological features described by Ladefoged (1971, 1975) are far from primitive linguistic phonetic parameters. This point is not made clear in any of these earlier works, and, indeed, as far as I am concerned, was not fully appreciated. I hope to make it clear now that linguistic phonetic descriptions require about 17 articulatory parameters, and about the same number of acoustic parameters. Other aspects of linguistic descriptions, such as accounts of sound patterns *within* languages, are undoubtedly best stated in terms of phonological features; and if these descriptions are to be explanatory, it is necessary for the features to relate to articulatory or auditory (or cortical) phenomena. But phonological features are certainly not sufficient for specifying the actual sounds of a language; nor are they in a one to one relationship with the minimal sets of parameters that are necessary and sufficient for this purpose.

We can get a first approximation to a minimal set of articulatory parameters by considering those that have been used in computer programs that synthesize speech. Some years ago Coker, Umeda and Browman (1973) showed that it is possible to use articulatory specifications to produce intelligible English. The input to their computer program was a string of phonetic segments that were changed by the program into ten articulatory parameters. In so far as the sounds produced were like English, these parameters were sufficient to specify the sounds involved.

When we consider a wider range of languages, we have to increase the number of parameters. An attempt to list such a set of articulatory parameters is given in Table 1. We will not consider all the items on this list. We will be able to see how the parametric approach differs from more traditional linguistic descriptions by considering only those parameters that specify the position of the tongue. Furthermore, the set of parameters listed is only a first approximation to those required. We do not yet know enough to be able to specify all and only the parameters required for linguistic contrasts. But I hope this list will be sufficient to give some impression of a possible set of phonetic parameters, and to show their relationship to more familiar phonological features.

Table 1. The necessary and sufficient articulatory parameters.

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

The first two parameters, front raising and back raising, are illustrated in Figure 1. They specify the position of the body of the tongue. The front raising parameter may be thought of as a movement from something like the position occurring in [o] to something like one in [i], as shown on the left of the figure. The back raising parameter specifies a movement from approximately [ɑ] to [u], as shown on the right. In each case the movement should really be thought of as a deviation from the reference position of the tongue, so that terms such as front raising-lowering and back raising-lowering might be more appropriate. Both parameters have been defined elsewhere (Harshman, Ladefoged and Goldstein 1977) in formal terms as deviations, in cm, of the position of the tongue of an average speaker from the position of that speaker's tongue in a reference position.

The tongue positions of all the non-rhotacized vowels of American English may be specified reasonably accurately in terms of these two parameters. For example, Figure 2 shows how the vowel [u] may be thought of as a certain amount of back raising, combined with a negative amount of front raising (i.e. a deviation below the reference line) which keeps the front of the tongue down and moves the body of the tongue further back. When the deviations corresponding to these parameters are added together the position of the tongue for [u] results.

Recent work (partly reported in Ladefoged, Harshman, Goldstein and Rice 1978) has indicated that these two parameters can be used to describe the tongue shapes of vowels in other languages reasonably well. Additional parameters that will be discussed later are necessary to account for some distinctive shapes. But the two parameters shown (or something very like them) will probably account for more of the variance found in the vowels of the languages of the world than any other two parameters for specifying tongue shapes.

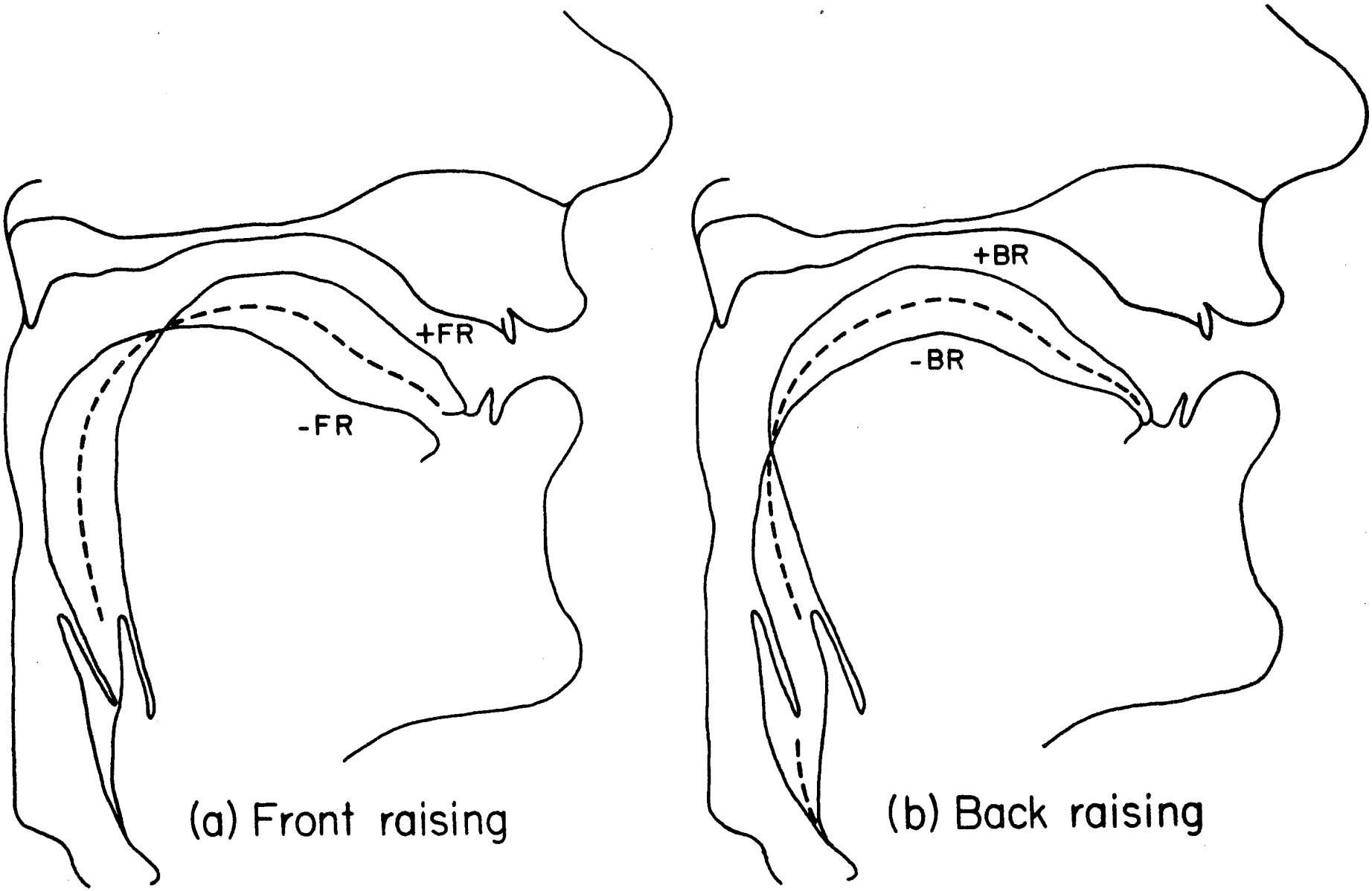


Figure 1. Positive and negative values of (a) the front raising parameter, and (b) the back raising parameter. The reference position is indicated by a dashed line.

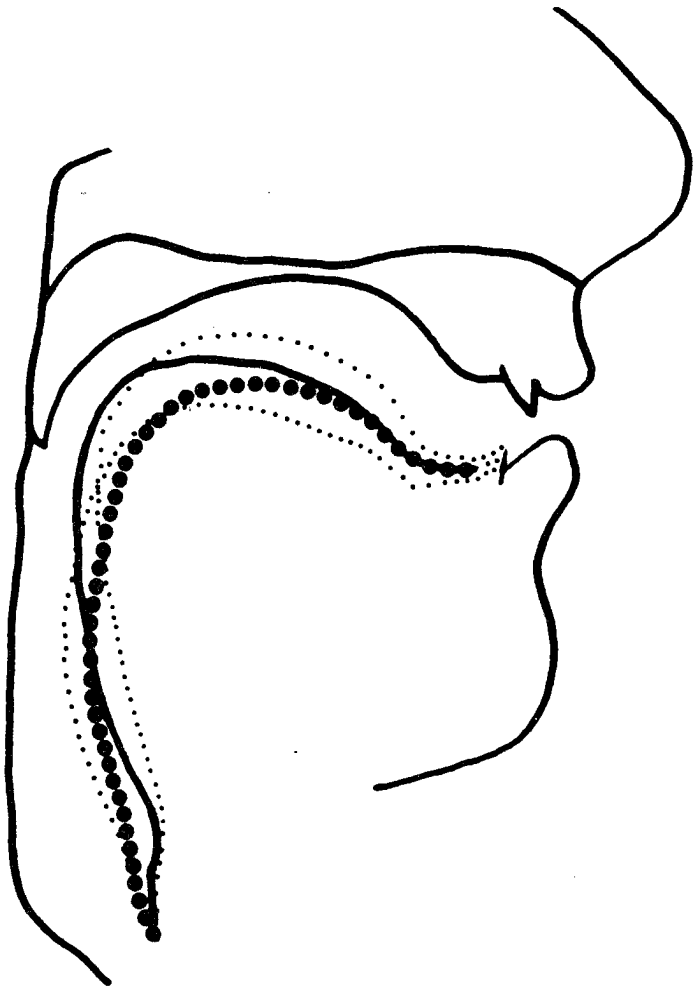


Figure 2. Reconstruction of the vowel /u/ as in "who" (solid line). The heavy dotted line indicates a reference position for the tongue, and the two light dotted lines indicate the deviations from the reference line of the two parameters that sum to give the deviation for /u/. The two dotted lines cross near the uvula. (For reasons of clarity the epiglottis is not shown.)

This claim instantly invites comparison with other systems of describing vowels, such as a more traditional description in terms of the position of the highest point of the tongue. The trouble with the traditional system is that it defines the location of only one point on the tongue, and there is no algorithm for describing the position of the rest of the tongue given just that information. Figure 3 shows two possible tongue positions that have the same highest point. Given only the location of the solid point there is no way of determining which of these (or many intermediate) shapes is being described. It may eventually be possible to use the highest point of the tongue to refer to unique, actually observed, tongue positions. But no one has demonstrated a method for doing this, and it is impossible to say how much of the variance among vowels could be accounted for in this way.

We must now consider whether descriptions of the body of the tongue in terms of front raising and back raising parameters are simply mathematical abstractions, or whether they can really help us to explain why vowels are as they are. It seems, in fact, as if they might well summarize some of the principal muscular forces involved. The tongue and mandible form a very complex system,with a wide variety of potential actions (Hardcastle 1976, Lieberman 1977). As may be seen from figure 4, the front raising-lowering parameter corresponds in great part of the actions of the genioglossus and opposing muscles such as the glossopharyngeus and other pharyngeal constrictors. The back raising-lowering parameter effectively summarises the opposing actions of the styloglossus and the hyoglossus. However, there are many possible compensatory actions of the jaw and the tongue muscles, and it is probably not too profitable to consider either of these parameters as simply specifying the action of a group of muscles. It seems more likely that these parameters (and perhaps the others that we will be discussing) describe higher level cortical control functions. That is, we may think of them as the underlying parameters that determine the synergistic actions that are required for the skilled motor movements that occur in speech.

But, we might well ask as linguists, does any of this have any explanatory power from our point of view? What is important to us is whether these parameters help us to account for phonological phenomena. This may be considered by seeing how they divide vowels into classes. Figure 5 (based on data in Ladefoged *et al*, shows the degree of front raising and back raising in a number of English vowels (mean values for five speakers). There is a very general similarity between the arrangement of the vowels in this figure and their location in a traditional vowel chart. The front raising parameter clearly separates front vowels from back vowels. But the back raising component, considered as a single physical scale, is not very useful in explanations of observed vowel patterns, or phonological rules for alternatives of vowels (although it does help explain articulatory similarities such as that between low back vowels and pharyngeals).

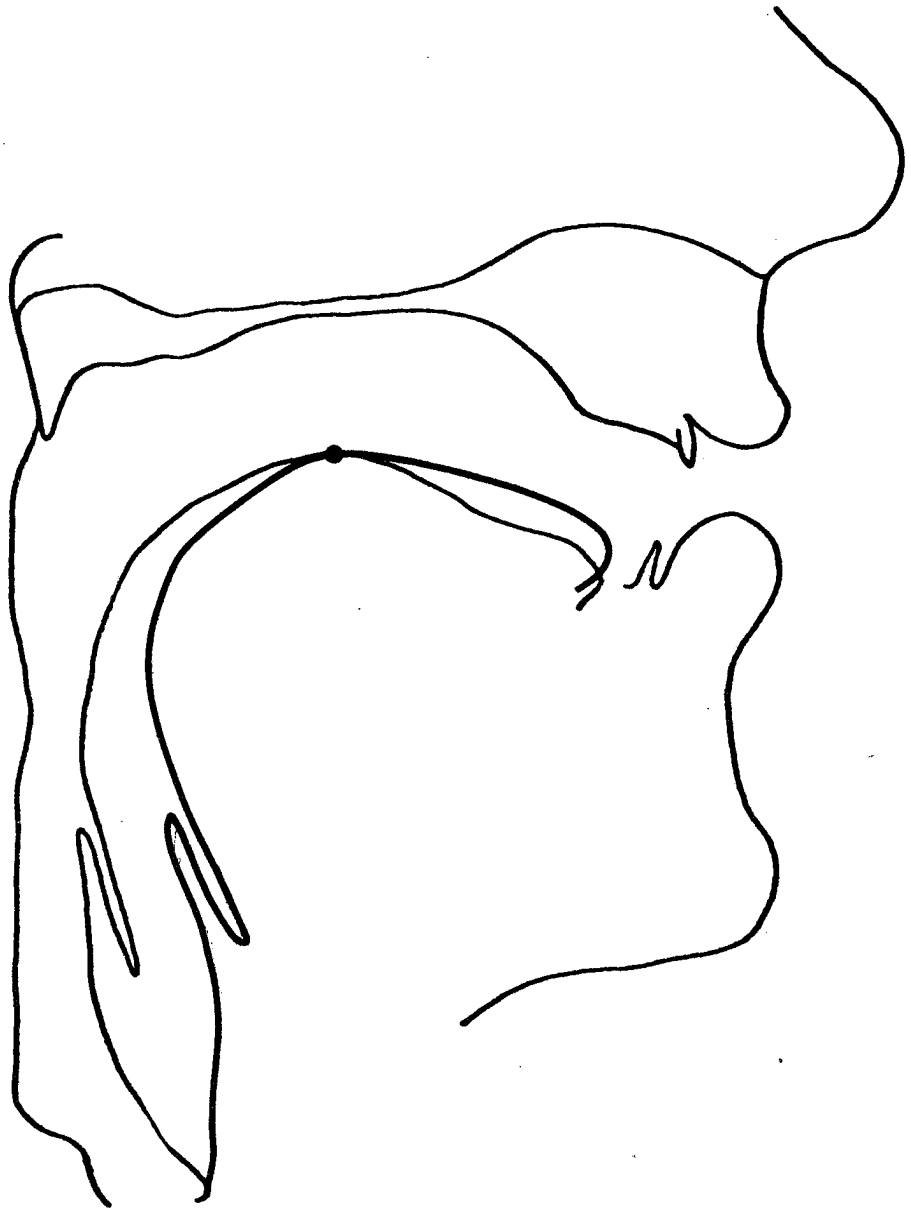


Figure 3. Two different tongue shapes with the same highest point of the tongue.

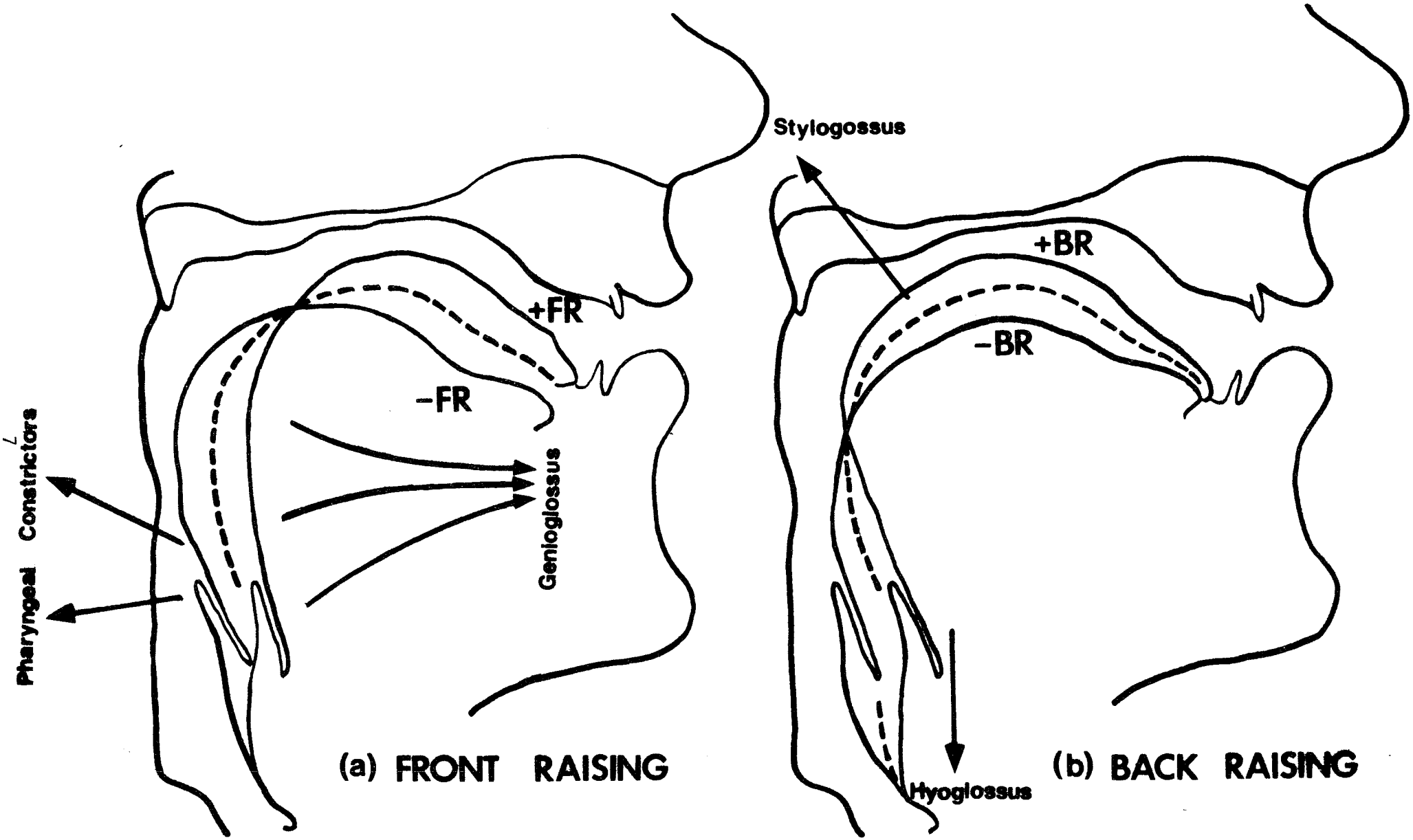


Figure 4. The principal muscular forces involved in (a) front raising and (b) back raising.

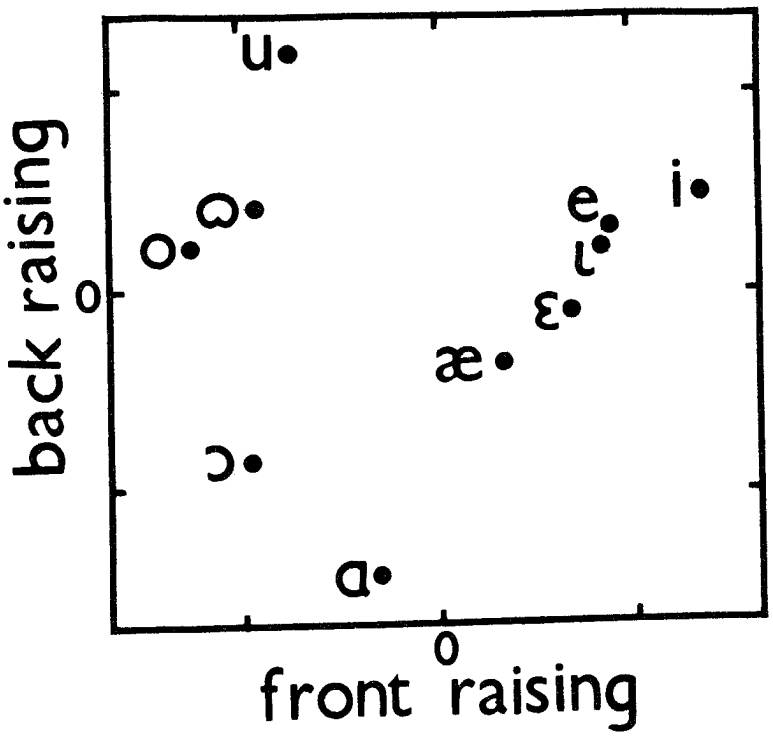


Figure 5. The degree of front raising and back raising in ten American English vowels (mean values of five speakers).

We have discussed these two articulatory parameters at some length in order to provide a good example of the lack of a match between linguistic phonetic descriptions and traditional phonological units. It is worth remembering that these articulatory parameters have been set up simply to account for linguistic differences among utterances, and in this sense are linguistic primes. This is a point to which we will return when we have discussed, somewhat more briefly, some of the other parameters in Table 1.

The next two parameters listed, tip raising and tip advancing, have a more straightforward function. They, too, can be defined in quantitative terms as deviations in cm from a reference position for an average speaker. Figure 6 shows the two dimensional movements of the tip of the tongue associated with retroflex, alveolar, and interdental positions. As a first approximation these movements may be considered as specifying variations in the position of the tongue that are independent of those specified by the parameters for the body of the tongue. Perhaps in years to come, when we hopefully have a larger body of good phonetic data, we will be able to take into account the correlation between movements of the tip of the tongue and those of the rest of the tongue. This correlation is particularly obvious in gestures such as sticking the tip of the tongue as far as possible out of the mouth -- a maneuver doctors use when they want the root of the tongue pulled forward out of the way so that they can view the larynx. But even in the comparatively small movements involved in speech there may be interactions. As Ohala (1974) has observed, alveolar and dental consonants seem to cause a lowering of the back of the tongue in some circumstances.

The phonological correlates of movements of the tip of the tongue are readily apparent. But it is worth noting that features such as Coronal (or Alveolar) can be defined only in terms of both tip raising and tip fronting. Given this particular set of articulatory parameters, there is no way that Coronal (or Alveolar) can be interpreted in terms of a single physical scale. Of course it would always be possible to say that linguists must include in their phonetic descriptions some additional, ad hoc, parameter, such as the distance to which the blade of the tongue is raised from its reference position. But this would be pointless because this parameter would be fully predictable from those that are already needed. The parameters being described are a necessary and sufficient set to account for all linguistic differences between utterances.

Let me emphasize that I am not suggesting that terms like Coronal or Alveolar should be replaced in phonological descriptions by terms like Tip raising and Tip advancing. When describing the sound patterns of languages we will want to refer to natural classes defined in terms of conventional phonological features; and these features must refer to observable phonetic phenomena. But I am advocating that when we are making a phonetic description of a language we do not do so by trying to interpret each feature in

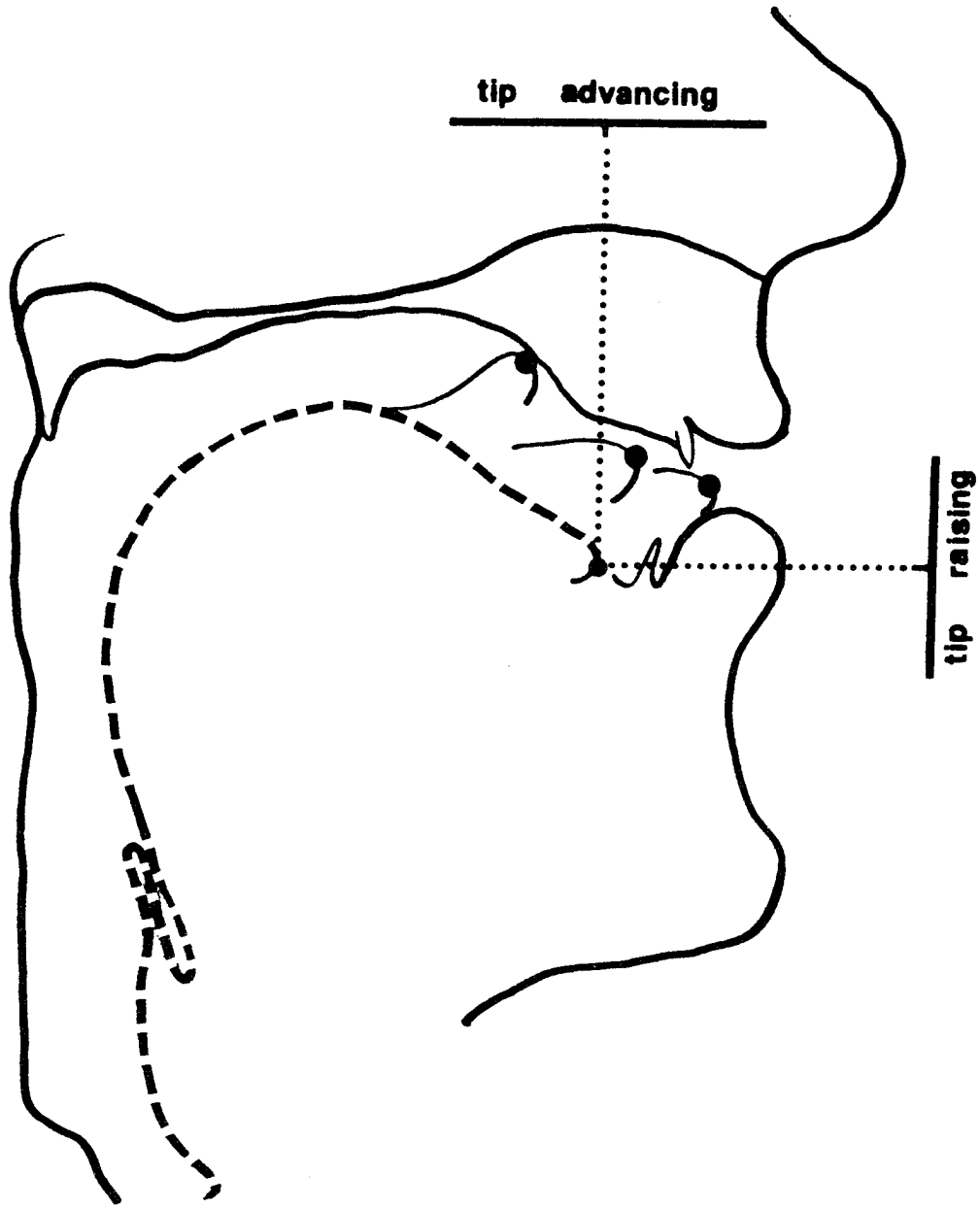


Figure 6. The variations in tip raising and tip advancing required for retroflex, alveolar,and interdental sounds.

terms of a single physical scale. We must be able to describe sound patterns in terms of phonological features. But we must also be able to map these features onto the basic linguistic phonetic parameters. The form of the mapping rules involved will be discussed later in this paper.

Similar points can be made with respect to other phonological features that are sometimes used for distinguishing among alveolar and dental sounds. The two degrees of freedom required for tongue tip movement can be combined with the parameters required for the specification of the body of the tongue to distinguish between apical and laminal sounds (or [∓ distributed] if this terminology is preferred). The apical and laminal categories are abstractions involving more than one of the physically definable parameters. They are not themselves part of the set of minimal phonetic parameters.

Additional parameters are required for specifying other aspects of tongue position. Among these are the variations in pharynx width (tongue root advancement) that occur in vowels (and perhaps in obstruents) and the bunching of the tongue that occurs in /r/ sounds, sometimes with and sometimes without a movement of the tip of the tongue. Figure 7 shows my best estimate at the moment of the deviations in tongue position that can be associated with each of these parameters. The diagram for tongue bunching should be regarded as especially tentative. It is based on an analysis of a very limited number of American English speakers saying words such as "heard." However, these data, together with the observations of Uldall (1958) and Delattre and Freeman (1968) and our own experience in synthesizing /r/ sounds from an articulatory model, all indicate that the shapes of the tongue that occur in these sounds cannot be produced without a contraction in the pharynx something like that shown.

Both pharynx width and tongue bunching do, in fact, correlate in a fairly simple way with phonological features. Because of lack of data, I cannot say much about tongue bunching other than that it seems to be in a monotonic relationship with the phonological feature Rhotacization (Ladefoged 1975). Pharynx width is worthy of further comment in that it correlates very highly with the feature Expanded as recently discussed by Lindau (1978), and it also offers an interesting insight into a problem that has troubled phonologists for some time. It has never been clear how one could give good phonetic definitions of overlapping phenomena. Thus in languages which have pairs of vowels such as [i,ɪ] and [e,ɛ], in which the distinction is said to be relative advancement of the tongue root, one always wants to know: relative to what? Using the parametric approach outlined here one can give an answer that will produce physical specifications for an average (or any other) speaker. The major aspects of the tongue position are the result of adding deviations from the reference position associated with the front raising and back raising parameters, and the variations in the position of the root of the tongue are additional deviations associated with the pharynx width parameter. Observed tongue shapes are the result of summing the actions of these three (and other) underlying parameters.

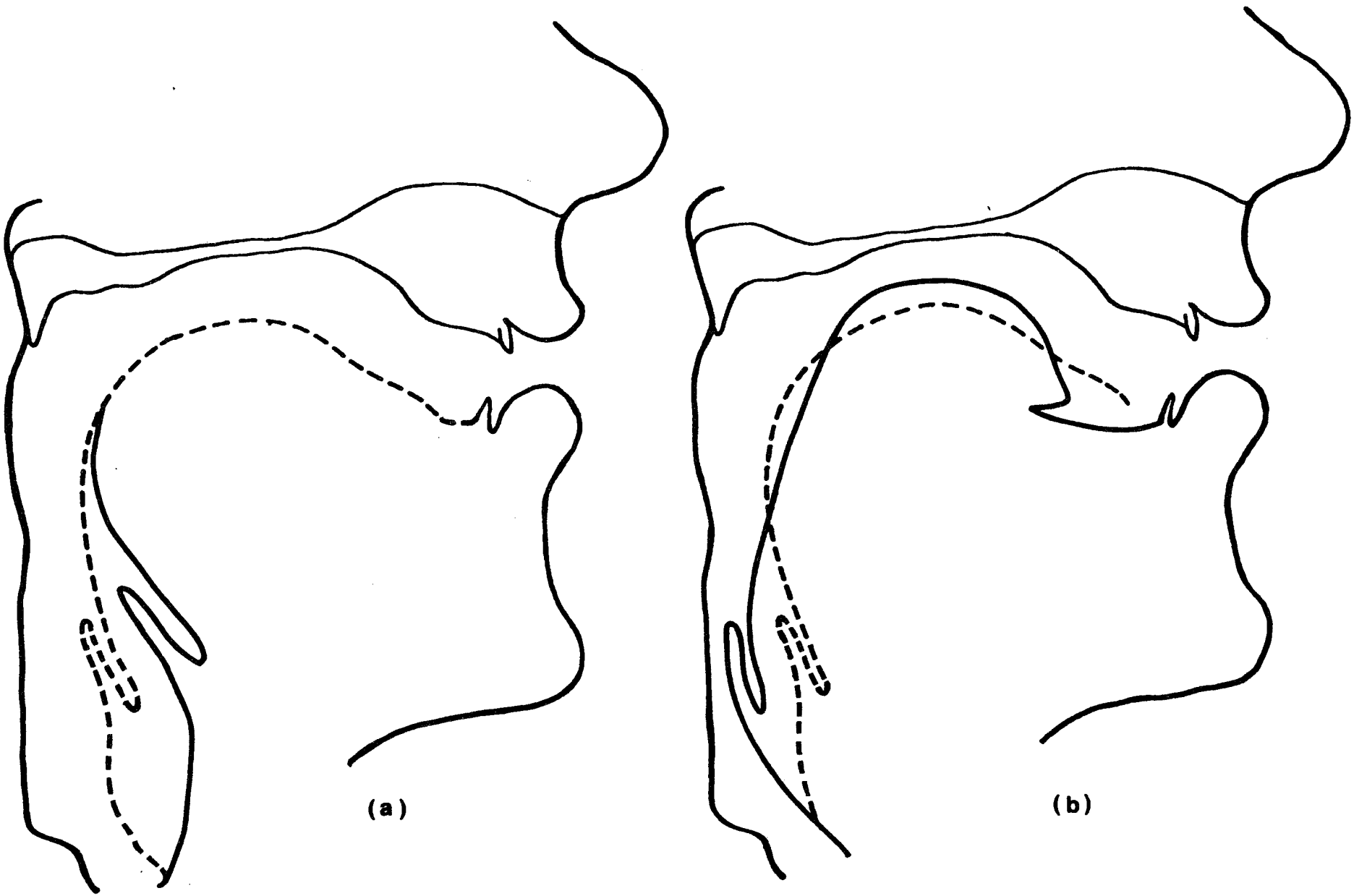


Figure 7. The movements of the tongue associated with (a) Pharynx width, and (b) tongue bunching

On tongue narrowing, the parameter associated with laterals, I have little to say except to express a hope that it will be possible to define it in such a way that it will account for dental, alveolar, retroflex, palatal, and velar laterals. I have no idea if this is possible. But it is only when we can give a formal account of what part of the tongue is narrowed that we will be able to give a really meaningful definition of the physical correlates of the feature Lateral.

Tongue hollowing I have included as an additional parameter to account for other variations in tongue shape in the coronal plane (the view from the front, as opposed to the more traditional sagittal view of the vocal organs). There are clear variations in the degree of hollowing and doming of the tongue in, for example, English [s] and [ʃ]. But I cannot give a quantitative description of these phenomena. Again let me emphasize that the parameters listed are illustrative of this approach, rather than definitive of what phoneticians have observed. The first two parameters were algorithmically derived from a limited set of English vowels (Harshman, Ladefoged and Goldstein 1977); but the remainder are simply my best estimates.

Many of the remaining parameters in Table 1. could have been used to demonstrate the relationship between phonological and phonetic units advanced here. But by now it should be clear that the necessary and sufficient set of articulatory parameters required for characterizing linguistic contrasts is not identical with the set of features required for characterizing phonological patterns. A similar point can be made by reference to the acoustic parameters of speech.

Authorities differ on the precise composition of the minimum set of acoustic parameters required for synthesizing human speech. The set of parameters used by the OVE 3 Speech Synthesizer (Liljencrants, 1968) is shown in Table 2. The adequacy of something like these parameters for describing speech has been demonstrated by Fant and his colleagues about 15 years ago.

Table 2. The necessary and sufficient acoustic parameters.

1. Voice source frequency
2. Voice source amplitude
3. Frequency of formant one
4. Frequency of formant two
5. Frequency of formant three
6. Bandwidth of formant one
7. Bandwidth of formant two
8. Bandwidth of formant three
9. Amplitude of nasal formant
10. Frequency of nasal formant
11. Amplitude of aspiration
12. Amplitude of fricative source
13. Frequency of lower fricative pole
14. Frequency of upper fricative pole
15. Relative amplitude of fricative poles

It might be objected that all the languages synthesized by Fant and his associates belonged to the IndoEuropean family, and we do not know if these parameters are sufficient to specify the complete range of possible linguistic contrasts. Personally I feel that the consonant combinations that occur in Russian are as tough a test for acoustic specification as anything found elsewhere. But in order to demonstrate that even so-called exotic sounds can be adequately specified in this way, I synthesized two non-IndoEuropean phrases (Ladefoged, in press). The first is part of a greeting exchange in Yoruba and contains a labial velar stop (Ladefoged 1968), [ɛkâr̀o.ó a dukṕɛ́] 'Good morning. Thank you.' The second is in Zulu, and contains two clicks, [g͡ǃ͡ba isiǂʰúmo] 'I saw an antelope.' There are some minor imperfections in the specification of these phrases, but there is nothing that indicates that any of these sounds cannot be specified in terms of appropriate time varying values of these or some similar set of acoustic parameters.

Some phonological features can be readily interpreted in terms of a single acoustic parameter. Thus vowel height is inversely proportional to the frequency of the first formant. But the majority of phonological features are in a many to many relationship with the minimal set of acoustic parameters, just as they are with the minimal set of articulatory parameters.

It is always possible to convert a specification of a particular speech sound in articulatory terms into a unique specification of that sound in acoustic terms. To a great extent it is also possible to reverse this process and convert an acoustic specification into a unique articulatory specification. There are a few cases where more than one articulatory configuration will produce a given sound. Atal and colleagues (1978) have shown that it is theoretically possible to produce some vowels with a variety of different vocal tract shapes. More importantly to us as linguists, Riordan (1977) has indicated how compensatory articulations may be involved in linguistic change. She found that in certain experimental circumstances, subjects will attempt to maintain a given vowel quality by using adjustments of the larynx position to compensate for different lip positions. We do not yet know how much these compensations are fundamental properties of language.

Work at UCLA (Ladefoged, 1979) indicates that the compensatory articulations that are possible are strictly limited. I cannot, for example, find any possible ways in which two distinct tongue positions will produce the same sound without there also being changes in the positions of the lips or the larynx. It is a matter of common experience that speakers usually use very similar articulations to produce a given sound. Everybody closes their lips to produce /p,b,m/ raises their tongue to the alveolar ridge to produce /t,d,n/, etc. In fact probably the only consonant in American English that does not have a unique articulatory specification corresponding to a particular acoustic specification is /r/, which may be produced with the tip of the tongue up or down, and with variation in the degree of bunching

of the tongue. Similarly, with vowels, "Whenever a speaker produces the vowel /i/ as in "heed" the body of the tongue is always raised up towards the hard palate. Whenever anyone produces the vowel /ɑ/ as in "father" the tongue is always low and somewhat retracted." (Ladefoged *et al* 1978)

This interconvertibility of articulatory and acoustic descriptions has recently been exploited by members of the UCLA Phonetics Lab (Ladefoged and Lindau, 1978). Our work on going from speech sounds back to vocal tract shapes has progressed to the stage where we can take a short sentence consisting of predominantly vocalic sounds and reconstruct a set of vocal tract shapes that might have produced this utterance. We cannot as yet reconstruct plausible vocal tract shapes corresponding to true consonants, simply because the appropriate algorithms have not yet been written. At the moment we can handle only such everyday utterances as "We owe you a yoyo" "How will you woo her away" and "We will weigh you." But there are no theoretical difficulties in going much further than this.

Since it is always possible to convert an articulatory description into an acoustic one and vice versa (though not necessarily uniquely), it might appear that either the acoustic parametric description or the articulatory description is redundant. We could say that one or the other comprises the minimal set of linguistic phonetic parameters, but not both since either set would allow us to make descriptively adequate statements. The articulatory parameters would also serve as a basis for explanatory statements by physiologists concerned with motor movements, and the acoustic parameters as a basis for explanatory statements by those interested in audition. But as linguists we will want to refer to both sets of parameters. Languages get to be the way they are because of the interplay between articulatory and acoustic (and other) factors. As we noted when discussing the two sets, some phonological features correlate in a simple way with parameters from one set, and others with the other. I used to think (Ladefoged 1971, 1975) that all but a few features which I termed cover features, could be defined in terms of either simple articulatory scales, or simple acoustic scales. It now appears to me that this is an oversimplification, and that very few features can be simply correlated with any of the minimal phonetic parameters.

Having considered the mapping of articulatory parameters onto acoustic parameters, and vice versa, we must now discuss the way in which systematic phonetic descriptions can be mapped onto parametric descriptions of either kind. As I have been emphasizing, I am in agreement with the standard view that phonologies should describe sound patterns by means of rules linking underlying forms with systematic phonetic descriptions. But the values assigned to the features at the systematic phonetic level are not full descriptions of the sounds. Taken as a set they are neither necessary nor sufficient to specify what it is that makes English sound like English rather than German. In order to map features onto articulatory or acoustic parameters we need something like a speech synthesis by rule program, which

would provide the necessary additional information. Thus the rules for mapping the three segments (each considered as a set of feature values) in [k^hæ?t] 'cat' have the general form

$$P_i = \alpha f([k^h]) + \beta f([æ]) + \gamma f([?t])$$

where P_i is the value for parameter i, f ([k^h]) is a function of the feature value of [k^h] (the particular allophone that occurs in 'cat'), f ([æ]) is a function of the feature values in the allophone [æ], and f([?t] of those in that allophone. The variables α, β, γ are time varying weighting functions corresponding to the degree of coarticulation that occurs in these circumstances. The functions for the allophones may be thought of as fairly straight forward look up tables. Thus to specify the position of the body of the tongue associated with [k^h] we may write:

(1) $\begin{bmatrix}+\text{velar}\\+\text{stop}\end{bmatrix} \rightarrow P_{\text{front raising}} = -1.0$

(2) $\begin{bmatrix}+\text{velar}\\+\text{stop}\end{bmatrix} \rightarrow P_{\text{back raising}} = +3.0$

Note that we cannot interpret [+velar] by itself, nor [+stop] by itself. The front raising and back raising parameters have to be determined by considering both these features together.

These mapping rules probably do not have any psychological reality. They are simply ways of relating one set of linguistic facts (phonological descriptions of the sound patterns) to another (phonetic descriptions of the sounds of one language as opposed to another). I think it most unlikely that speakers organize the production or perception of spoken language in terms of segments and features. The phonological patterns that linguists observe are abstract properties of the social institution we call language. But the articulatory and acoustic parameters are defining constraints of human behavior.

At this stage perhaps more evidence is needed as to why either set of parameters is the *minimal* set of linguistic units. I have simply asserted that these parameters are necessary and sufficient for phonetically characterizing linguistic contrasts. That they are necessary is evident from attempts to synthesize speech; if we omit the values for one of them we will be unable to produce certain contrasts. That they are sufficient is a harder claim to justify; but it can easily be disproved by finding counterevidence. At the moment there is no reason to believe that such evidence is likely to become available. All the recently described phonological contrasts seem to be new combinations of previously known possibilities rather than totally new phenomena. For example the velar laterals in Melpa and Mid-Waghi exemplified in Table 3 (Ladefoged, Disner and Cochran, 1977) involve no new parameters. Similarly the bilabial trills in Kele and Titan which are examplified in Table 4 (ibid) can be described in terms of the lip and vocal cord parameters in Table 1, or the formant and larynx source parameters in Table 2. The different voice qualities in Mpi (Harris and Ladefoged, forthcoming) illustrated in Table 5 can be accounted for in terms

Table 3. Contrasts between dental, alveolar, and velar laterals

Melpa (Eastern Highlands, Papua New Guinea)

	Medial		Final	
Dental	kia̪l̪t̪im	'fingernail'	waɬ̥	'knitted bag'
Alveolar	lola	'to speak improperly'	baɬ̥	'apron'
Velar	paga	'fence'	rag̊	'two'

Mid-Waghi (Eastern Highlands, Papua New Guinea)

a̪l̪aa̪l̪a 'again and again' alala 'to speak improperly' agage 'dizzy'

Table 4. Contrasts between bilabial and lingual trills
Kele (Austronesian, Papua New Guinea)

Bilabial	mɃin	'vagina'	mɃulim	'your face'	mɃɛŋkei	(a fruit)
Lingual	nruwin	'bone'	nrilɛŋ	'song'	nrikei	'leg'

Titan (Austronesian, Papua New Guinea)

Bilabial	mɃulei	'rat'	mɃutukei	'wooden plate'
Lingual	ndruli	'sandpiper'	ndrake?in	'girls'

Table 5. Contrasts between the six tones and between the plain and laryngealized vowels in Mpi (Southern Lolo branch, Tibeto-Burman, Northern Thailand).

	1. ˩	2. ˧˩	3. ˧	4. ˧˧	5. ˧˥	6. ˥
Plain	si	si	si	si	si	si
	'to be putrid'	'blood'	'to roll'	(a color)	'to die'	'four'
Laryngealized	sḭ	sḭ	sḭ	sḭ	sḭ	sḭ
	'to be dried up'	'seven'	'to smoke'	(classifier)	(man's name)	(man's name)

of the glottal parameters listed in Table 1. In this case the acoustic parameters listed in Table 2 may have to be expanded to allow for differences in glottal pulse shape that are apparently controllable in Mpi. But Laver (1977) has shown that a very remarkable range of phonation types can be synthesized in terms of comparatively simple combinations of acoustic parameters.

There is potentially a far larger range of speech sounds that could be contrastive. Several other phonological contrasts might occur in as yet undescribed languages, some of which might be very difficult to describe in terms of the present parameters. There are, for example, no known languages that use lateral movements of the tongue between the lips, but any child can make a noise of this type. To the best of my knowledge there is no language that uses buccal fricatives in which movements of the cheeks produce an egressive airstream. These sounds are also fairly simple to learn. All such sounds might be considered to be part of the "phonetic capabilities of man" (Chomsky and Halle 1968). It would be completely improper to disregard them simply because they have not yet been observed in some language. However, until I am proved wrong by events, I will contend that they need not be characterizable in terms of either set of parameters, on the grounds that they are too hard to integrate into a spoken language.

I am of course, aware that this is a vague pronouncement, a sort of hand waving indicating that something might be done to make the theory being proposed more testable. In reality, the validity of these parameters,(like much of linguistics) is not a scientific notion that is dependent on an empirically testable hypothesis. As Abercrombie (1956) has pointed out, tests that involve knowing all present, past and future languages are obviously pseudo-procedures. A theory concerning the phonetic capabilities of man is inevitably simply a description of the known data on the basis of which one can provide speculative (but never scientifically proven) explanations that predict what is likely to be observed in the future. Linguists, like any other group of scientists/observers, should seek explanations for the regularities they observe, but should not be worried if their explanations are merely predictive of future observations and not otherwise testable. However, this view does not absolve us from the responsibility of, whenever possible, expressing our observations in terms of numbers that can be shown to be valid, reliable and significant. Linguistic phonetic descriptions can do this appropriately by reference to the parameters in Tables 1 and 2.

Having spent much of this paper discussing contrasts within individual languages, we must now consider how to describe measureable phonetic differences between languages. The sounds of one language may differ from those of another because of the phonetic value of the segments. These differences are as much linguistic properties of the languages as are the differences in the sound patterns that are often more fully described. As linguists, we tend to get so involved with describing the phonology of, say, English or Danish, that we forget to point out that many of the sounds of

English are not the same as the similarly specified sounds of Danish. Thus even the monumental work *The Sound Pattern of English* (Chomsky and Halle 1968) is only a description of the patterns and not of the sounds. The theory may make it possible to give precise descriptions of the sounds of English. But the fact remains that *The Sound Pattern of English* does not tell us all that we need to know about the phonetic properties of a vowel that is specified as, for example, [+high, -low, -back]. We cannot tell if it sounds the same as a vowel that may be similarly specified in a description of Danish. As phoneticians have long known, /i/ in English is not the same as /i/ in Danish, and a complete linguistic description of each of these languages must make this evident.

The inadequacy of current phonological theories becomes more apparent when we consider sounds such as the velar ejectives in Hausa and Navaho. These consonants may be given the same label, and written with the same symbol, [k'], in a phonetic transcription. But they do not sound the same. If a Navaho speaker used a Hausa velar ejective while speaking Navaho it would sound like Navaho spoken with a foreign accent. It is very difficult to describe differences of this kind in terms of phonological features. But if there is a noticeable difference between two sounds in different languages such that either of them would sound foreign if it were used in the other language, then this difference is part of the linguistic facts of each language.

I will now consider two cases in which there are measureable phonetic differences between languages that should be evident from full descriptions of each language. In neither case can the differences between languages be taken into account by some notion of variation in the basis of articulation. In fact the whole concept of base of articulation seems to me to be invariably inadequate for discussing differences between languages. I know of no quantified differences between languages that can be handled in this way. In every case, when giving a precise account of what makes a particular language sound the way it does, it is necessary to describe the phonetic properties of individual segments.

The first set of data I will use to illustrate this point comes from work by Disner (1978). She has compared the vowels of Germanic languages, and has been able to substantiate traditional phoneticians' auditory judgments of the phonetic differences among these languages. For example, her plot of the formant frequencies of some of the long vowels of Danish is given in Figure 8. Each ellipse is centered at the mean of the reported formant frequencies for the vowel (Fischer-Jørgensen 1972), and has axes with lengths of two standard deviations. For comparison, the locations of four of the vowels of English are shown by dashed lines. The frequency values have been plotted on scales such that distances between points reflect perceptual distances. It is obvious that Danish /i/ is higher than English /i/. Moreover the four front unrounded vowels of Danish are unevenly spaced, three of them being much higher than their English counterprats. We

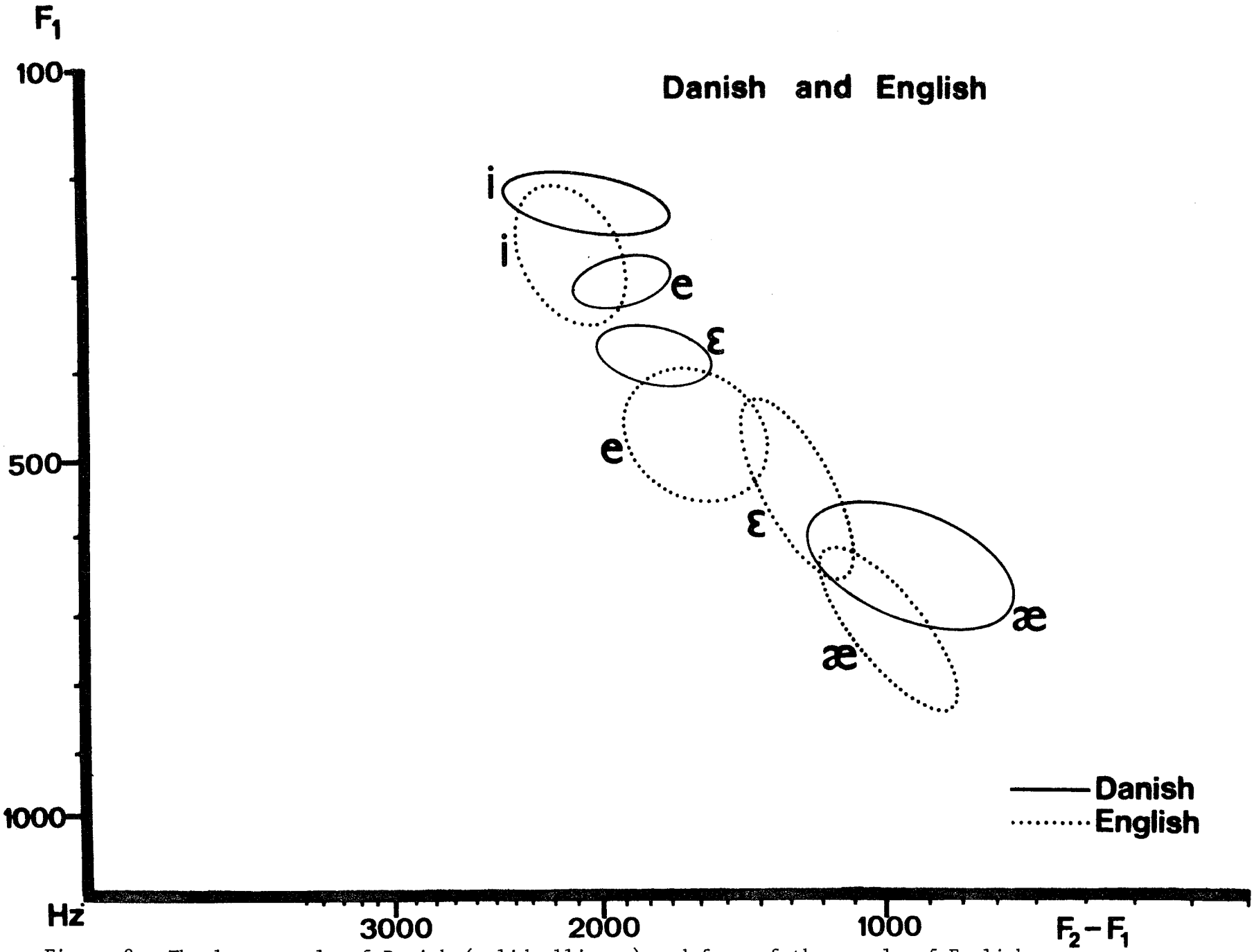


Figure 8. The long vowels of Danish (solid ellipses) and four of the vowels of English (dotted ellipses). From Disner (1978).

cannot say that Danish has a higher basis of articulation than English, because there is no uniformity to the difference between the two languages. Each of the Danish vowels is higher than its English counterpart, but the difference varies for each pair of vowels. We need specific descriptions of each vowel in each language in order to show how the vowels of the one language are phonetically distinct from similarly specified vowels in the other.

In this particular case the phonetic differences between the two languages can be expressed in terms of scalar values of features such as Height (or High and Low) that have simple acoustic correlates. But the phonological features that have been used to describe sound patterns within languges are in many cases not sufficient to account for linguistically significant differences between languages. Thus both Kalabari and Hausa, two languages that are spoken in Nigeria, have voiced glottalized bilabial and alveolar stops -- sounds that are usually transcribed as /ɓ,ɗ/. Figure 9 shows spectrograms of words containing these sounds preceded and followed by low vowels. There is a considerable difference between the two languages. In the Hausa words in the upper row, the preceding vowel is marked by irregular vibrations of the vocal cords, and there is at best laryngealized voicing throughout the closure. But in the Kalabari words in the lower row the implosive sounds are fully voiced throughout the closure and there is no tendency towards creaky voice or laryngealization.

I have investigated recordings of a number of speakers of each of these languages, and there is no doubt that this is a reliable, quantifiable, significant difference between these two languages. This difference can be described in terms of the parameters listed in Tables 1 and 2. But it cannot be handled in terms of the features suggested by Chomsky and Halle (1968). It might be possible to use the features suggested by Halle and Stevens (1971) which involve laryngeal parameters very similar to those listed in Table 1. But these features, like those in Table 1, do not enable us to categorize sounds into phonologically appropriate natural classes. For example English sounds that differ in voicing have to be distinguished in terms of two separate features, which Halle and Stevens call Stiff and Constricted. Again we see that the features that are necessary and sufficient for describing the phonetic properties of languages are not in a one to one relation with the features required for phonological descriptions.

The differences between Kalabari and Hausa implosives is not known to be contrastive within any one language. But there seems to be no principled reason why it should not be. It is perfectly audible, even to speakers of languages that do not have any of these sounds. The same is true of several other differences between languages. There are reported differences in the kind of lip rounding that occurs in French and German, for example. The sibilants of Swedish and Polish may also differ. And the tap r-sounds in Hausa and Malayalam may involve different degrees of lowering of the frequency of the third formant. There is no doubt that speakers

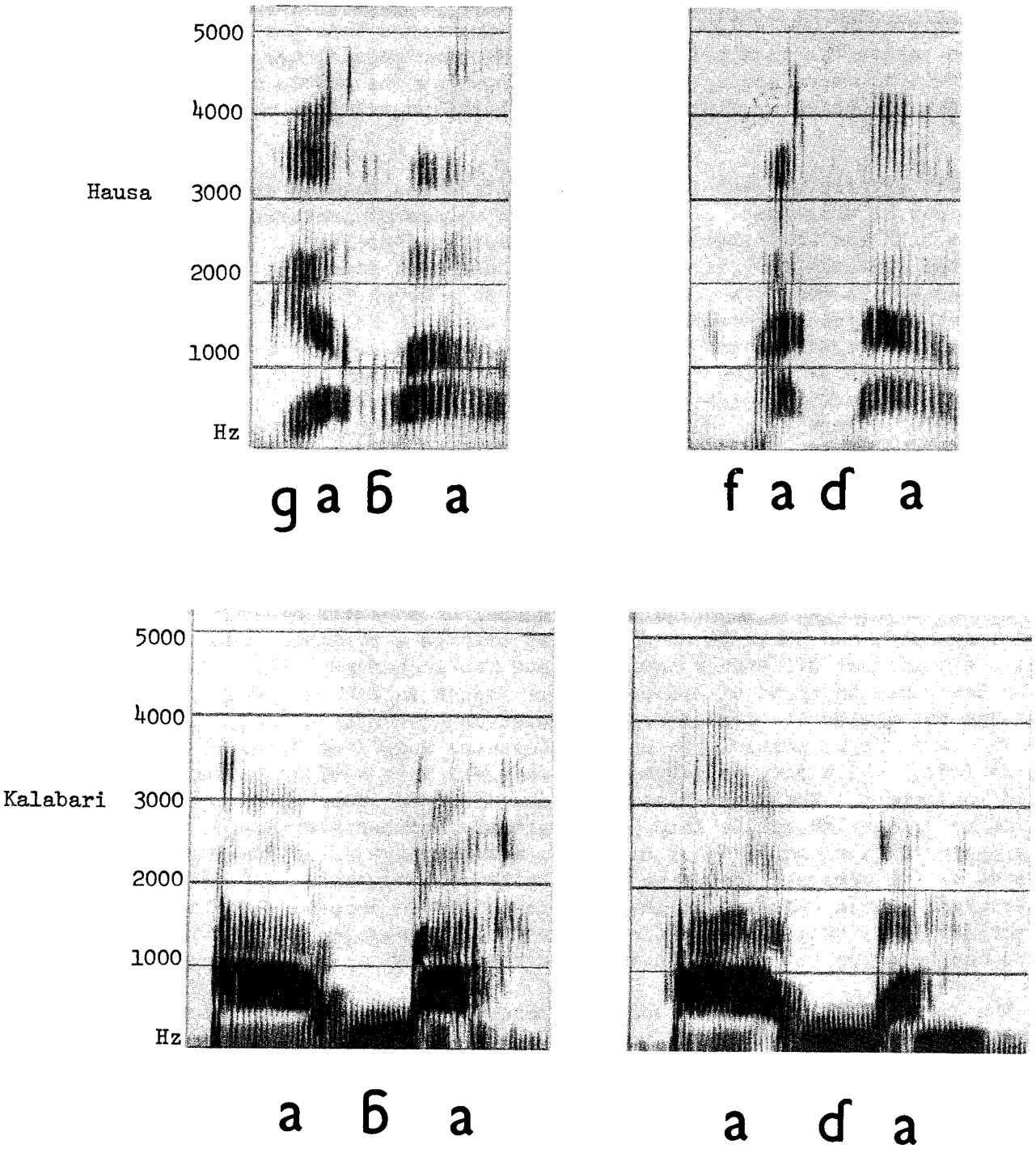


Figure 9. Voiced bilabial and alveolar implosives in Hausa /gáɓàa/ 'joint' /fáɗàa/ 'quarrel' and Kalabari /aɓa/ 'to kill her' /aɗa/'her father'

can make and listeners can hear at least some of these differences completely reliably. Therefore this degree of phonetic detail must be included in linguistic phonetic descriptions of languages.

In summary, I have tried to show that that the fundamental linguistic phonetic constraints are sets of articulatory or acoustic parameters. Each set is a necessary and sufficient set of parameters that will account for all possible linguistic phonetic properties. Descriptions in terms of one set can be converted into descriptions in terms of the other. Descriptions of phonological patterns in languages involve features which are quite distinct from the phonetic parameters. Moreover they cannot account for many of the phonetic differences between languages. At some abstract levels languages may be organized partly in terms of phonological features. But we must always remember that languages are complex properties of human societies, not of individual brains. Individuals producing and interpreting linguistic events probably use something like the parameters in Tables 1 and 2.

Acknowledgements

Many of the UCLA Phonetics Lab group have hacked critically at drafts of this paper in a series of lab meetings. Their vociferous comments have been a great help. I also received several useful comments from David Isenberg.

References

Abercrombie, D. 1956. Pseudo-procedures in linguistics. Lecture given at the University of Edinburgh.

Atal, B.; J. J. Chang; M. V. Mathews; and J. W. Tukey. 1978. Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. Journal of the Acoustical Society of America 63.1535-56.

Chomsky, N., and M. Halle. 1968. Sound pattern of English. New York: Harper and Row.

Coker, C.; N. Umeda; and C. Browman. 1973. Automatic synthesis from ordinary English text. IEEE Trans. Electracoust. AU-21.293-8.

Delattre, P., and D. C. Freeman. 1968. A dialect study of American r's by x-ray. Linguistics 44.29-68.

Disner, S. 1978. Vowels in Germanic languages. (UCLA Working Papers in Phonetics, 40). Los Angeles.

Fischer-Jørgensen, E. 1972. Formant frequencies of long and short Danish vowels. Studies for Einar Haugen, ed. by E. S. Firchow et al.,189-213. The Hague: Mouton.

Halle, M., and K. N. Stevens. 1971 A note on laryngeal features. Quarterly Progress Report, MIT Research Laboratory of Electronics, 101.198-213.

Harshman, R.; P. Ladefoged; and L. Goldstein. 1977. Factor analysis of tongue shapes. Journal of the Acoustical Society of America 62.693-707.

Ladefoged, P. 1968. A phonetic study of West African languages. Revised ed. Cambridge: Cambridge University Press.

——. 1971. Preliminaries to linguistic phonetics. Chicago: University of Chicago Press.

—. 1975. A course in phonetics. New York:Harcourt,Brace and Jovanovich.
—. 1979. Articulatory parameters. Proceedings of the IXth International Congress of Phonetic Sciences, Copenhagen.
—. In press. The phonetic specification of languages of the world. Proceedings of the VIIIth International Congress of Phonetic Sciences, Leeds.
—; A. Cochran; and S. Disner. 1977. Laterals and trills. Journal of the International Phonetic Association 7.46-54.
—; R. Harshman; L. Goldstein; and L. Rice. 1978. Generating vocal tract shapes from formant frequencies. Journal of the Acoustical Society of America 64.1027-35.
—, and M. Lindau. 1978. Prediction of vocal tract shapes in utterances. Journal of the Acoustical Society of America 64.S41.
Laver, J. 1977. Voice quality. Paper presented to the 1st International Phonetic Sciences Congress, Miami Beach.
Lieberman, P. 1977. Speech physiology and acoustic phonetics: an introduction. New York: Macmillan.
Liljencrants, J. 1968. The Ove III speech synthesizer. IEEE Trans. Electroacoust. AU-16.137-40.
Lindau, M. 1978. Vowel features. Language 54.541-63.
Ohala, J. 1974. Phonetic explanation in phonology. Papers from the Parasession on Natural Phonology,251-75. Chicago: Chicago Linguistic Society.
Riordan, C. 1977. Control of vocal-tract length in speech. Journal of the Acoustical Society of America 62.998-1002.
Uldall, E. T. 1958. American 'molar' r and 'flapped' r. Revista do Laboratorio de Fonetica Experimental, Coimbra 4.103-6.

Articulatory Parameters

Peter Ladefoged

[Paper to be presented at the 9th International Congress of Phonetic Sciences, Copenhagen, Denmark, August 1979]

The main report for this session gives an excellent summary of recent research on speech production. I would like to try to summarise this summary by listing and discussing the articulatory parameters that need to be controlled in a model of the speech production process. Obviously this could be done at various levels of generality. For example, one could choose to model the various muscular forces acting on the tongue, as suggested by Fujimura and Kakita (1978), or one could model the results of those forces as described by Harshman et al. (1977). Similarly one could specify the gross respiratory movements as Ohala (1974) has done, or more simply the variations in subglottal pressure that result from those movements. On another dimension of generality, one could try to describe just those articulatory parameters required for a particular language, or the larger set that would produce all possible linguistic differences, or even those that would go still further and allow one to distinguish all the personal characteristics of individual speakers.

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

The parameters listed may be thought of as corresponding to what is controlled rather than to movements of anatomical structures such as the jaw or the ribcage. This is a somewhat controversial point in that Lindblom and Sundberg (1971) have proposed that it is more appropriate to model tongue movements with respect to a moving mandible, rather than simply modeling the vocal tract shapes that result from these tongue movements. But it seems to me that if one is trying to state the parameters that are used in controlling articulatory actions,then Lindblom's own work (Lindblom et al., 1978) shows that speakers may rely on a great deal of compensation between movements of the jaw and those of the tongue. What they control are the vocal tract shapes, i.e.the relative magnitudes of the cross-sectional areas of the mouth and pharynx. The underlying parameters may therefore be as shown in table 1.

Table 1. A necessary and sufficient set of articulatory parameters.

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

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

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

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

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

It is obviously of interest to phoneticians to compare descriptions in terms of front raising and back raising with more traditional descriptions in terms of the highest point of the tongue, but unfortunately this cannot be done at the moment. The problem with these traditional descriptions is that no one has as yet shown how to interpret them unambiguously. Given the height and degree of backness of the highest point of the tongue (and given that all the other parameters such as pharynx width have neutral values) it is not yet known how (or even if) the position of the tongue as a whole may be described.

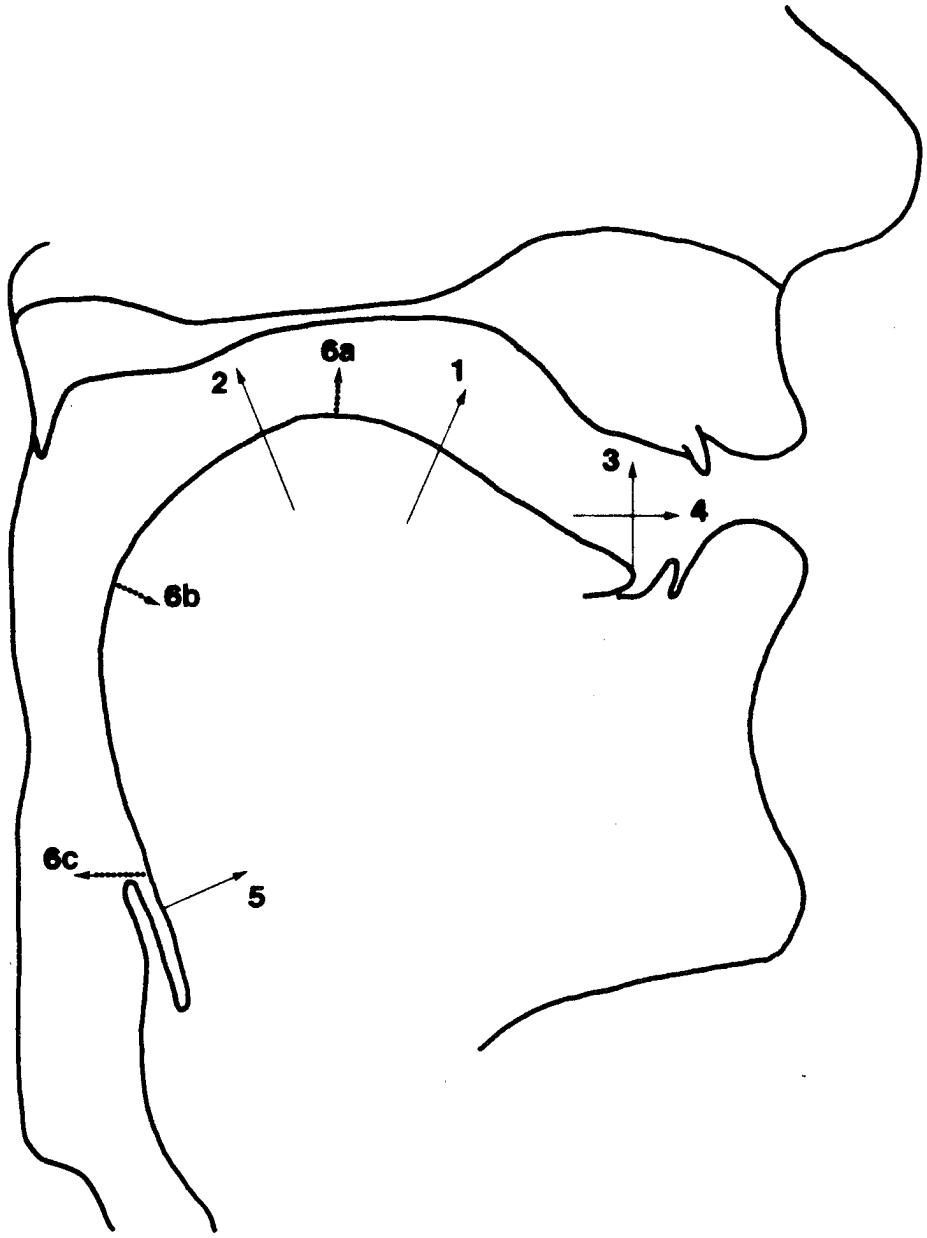


Figure 1. The movements principal portions of the tongue associated with the first 6 parameters in Table 1.

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

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

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

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

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

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

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

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

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

There are probably only two major ways in which variations in one parameter may lead to no change in the speech sound produced because they are offset by variations in another parameter. The first is the use of larynx lowering to offset decreases in lip rounding (Atal et al 1977, Riordan 1977). The second is the use of increased respiratory power (lung volume decrement) to offset decreases in the stretching of the vocal cords (glottal length). There may also be variations among the three lip parameters that can be used to compensate for one another. But the data of Atal et al (1977) on parameterized tongue shapes,and our own similar data, indicate that there are no cases in which a given sound can be produced with the same lip and larynx position, but with two different tongue shapes, as long as the tongue shape is characterized by only two parameters. There are

well known cases involving additional parameters, such as American English rhotacized vowels that may be produced in two different ways (Uldall 1958). There may also be variations in pharynx width that can compensate at least in part for variations in front raising and back raising to produce similar tongue shapes in vowels. But apart from the case of rhotacized vowels, I doubt that there are two distinct tongue shapes that produce the same sound.

The 16 parameters listed are hypothesized to be a necessary and sufficient set for linguistic phonetic specifications. Some of them are far from fully defined, but they are all susceptible of precise numerical specification. They are potentially the things that are controlled in speech production. As MacNeilage indicates, we do not yet know whether speech production involved specifying a sequence of targets or whether some form of action theory specification is preferable. The parametric approach outlined above is equally applicable in either case. Very tentatively, Table 1 is offered as a summary of what we use in speech production.

References

Atal, B., J.J. Change, M.V. Mathews and J. W. Tukey (1978) "Inversion of articulatory to acoustic transformation by a computer sorting technique," JASA 63, 1535-1555.

Catford, J.C. (1977) Fundamental Problems in Phonetics. Bloomington: Indiana University Press.

Fujimura, O. and Y. Kakita (1978) "Remarks on quantitative description of the lingual articulation," in Frontiers of Speech Communication Research. B. Lindblom and S. Ohman (eds.) London: Academic Press.

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

Ladefoged, P., J. DeClerk, M. Lindau, and G.A. Papçun (1972) "An auditory-motor theory of speech production," UCLA Working Papers in Phonetics 22, 48-75.

Ladefoged, P., R. Harshman, L. Goldstein, and D.L. Rice (1978) "Generation of vocal tract shapes from formant frequencies," JASA 64.4. 1027-1035.

Ladefoged, P., and R. Harshman (1979) "Formant frequencies and movements of the tongue". Frontiers of Speech Communication Research. B. Lindblom and S. Ohman (eds.).

Lindblom, R., J. Lubker and T. Gay (1978) "Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation," JPh (in press).

Lindblom, B.E.F. and J.E.F. Sundberg (1971) "Acoustical consequences of lip, tongue,jaw and larynx movement," JASA 50, 1166-1179.

Lindau, M. (1979) "The feature Expanded," JPh (in press).

Maddieson, I. (personal communication) "Phonological inventories of the languages of the world."

Ohala, J. (1974) "A mathematical model of speech aerodynamics," Proc. Speech Comm. Sem. Stockholm, 65-72, Uppsala: Almqvist and Wiksell.

Ohala, J. (1974a) "Phonetic explanation in phonology," Papers from the parasession on natural phonology, 251-71. Chicago: Chicago Linguistic Society.

Riordan, C.J. (1977) "Control of vocal-tract length in speech," JASA 62, 998-1002.

Uldall, E.T. (1958) "American 'molar' r and 'flapped' r," Revista do Laboratorio de Fonetica Experimental, Coimbra 4. 103-106.

Prediction of vocal tract shapes in utterances

Peter Ladefoged and Mona Lindau

[Revised version of paper presented at the 96th Meeting of

the Acoustical Society of America, Honolulu, Hawaii, November 1978]

Previous reports (Harshman, Ladefoged and Goldstein, 1977; Ladefoged, Harshman, Goldstein, and Rice, 1978; Ladefoged and Harshman, forthcoming) have described a system for generating vocal tract shapes from acoustic data for American English vowels in isolation. This is a progress report on the same system, now used for generating continuous vocal tract shapes from connected speech.

We recorded three speakers saying utterances that contained mainly vowels and vowel-like sounds. These were everyday conversational phrases like "We owe you a yoyo", "Where were you a year ago?", and "We may go away now". The first three formant frequencies were determined using a computerized LPC system. In addition we also used spectrograms of these utterances to correct and hand-edit the LPC files. It was not our purpose to develop a formant tracking system as a major part of our work.

Possible positions of the body of the tongue in these utterances were determined from the two parameters, Front Raising and Back Raising, which have been shown by Harshman et al. (1977) to specify the shape of the whole tongue. The algorithms for relating formant frequencies to these two parameters were determined by multiple regressions as described in Ladefoged et al. (1978).

A third tongue shape parameter has been added to the previous system to include the rather curious type of bunching that many speakers use in the r-colored vowels of American English. Figure 1 shows a tracing of [əɹ] in *herd*, spoken by an American speaker from cineradiographic data. This type of r-bunching cannot be generated by any combination of Front Raising and Back Raising. The r-bunching tongue parameter is associated with a low third formant (below 1900 Hz) and a relatively short distance between the second and third formants.

In previous work the distance between the upper and lower lip was predicted from a combination of the first and second formant frequencies. Departing from this, the distance between upper and lower lip is now generated as a straightforward function of the frequency of the first formant, as in (1).

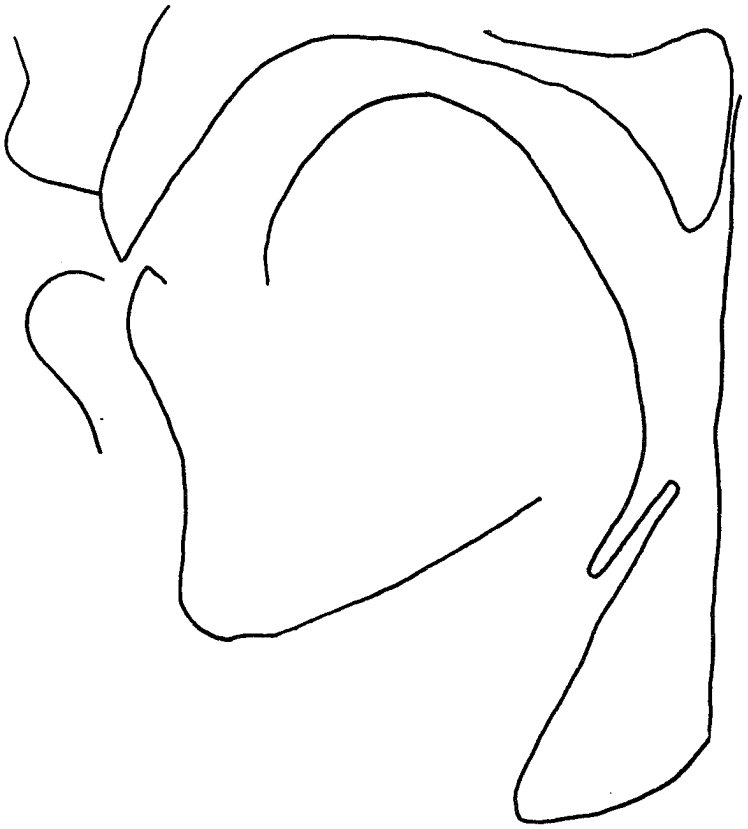


Figure 1. Tracing of the r-colored vowel in *herd*.

$$H = 0.0044\ F1 - 0.5 \qquad (1)$$

where H is the distance between the upper and lower lip.

The width of the lips is a function of the sum of F1 and F2.

$$W = 0.00355\ (F1 + F2) - 3$$

where W is the distance between the corners of the lips.

These five parameters, Front Raising, Back Raising, Tongue Bunching, Lip Height, and Lip Width, were used to generate a display of the vocal tract of the kind seen in Figure 2. A traditional sagittal view of the position of the tongue and lips was generated on the left of the computer screen. A frontal display of the lips was generated on the right of the screen.

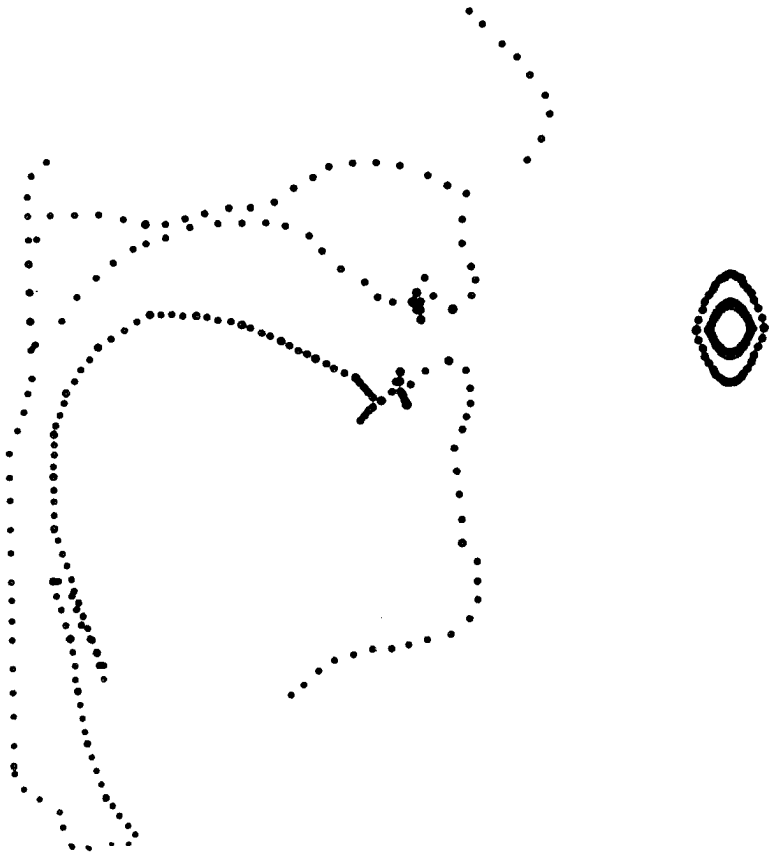


Figure 2. Computer-generated display of the vocal tract and lips.

The positions of the velum and the tip of the tongue can also be varied in the computer generated display, but we have not yet implemented any algorithms for determining velic position or tongue tip position from acoustic data. The nasals in "We may go away now" were done by hand.

Sequences of positions of the vocal organs were generated from formant frequencies for all the vowels and vowel-like sounds in the recorded utterances. These algorithms are quite successful in producing plausible pictures.

This system is inadequate for generating obstruents, i.e. plosives and fricatives, such as the [g] in *ago*. Figure 3 shows that the available algorithms will produce some narrowing of the vocal tract in the correct region for [g], but the amount of raising of the tongue body towards the velum is not enough. For these types of sounds there is a different relationship between acoustic structures and articulations than for vowels.

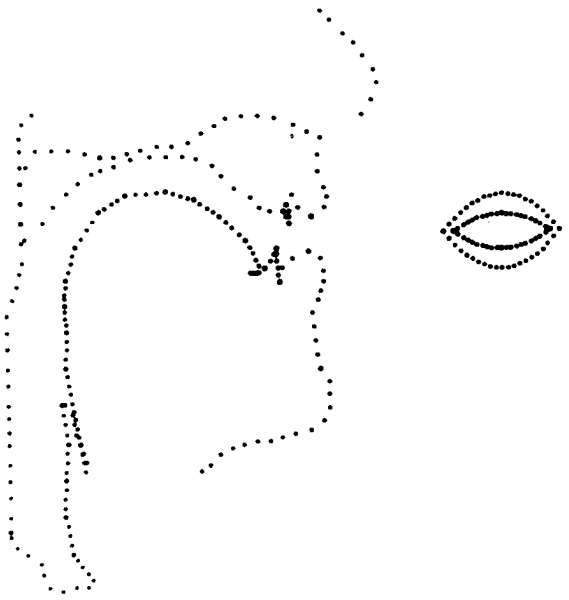


Figure 3. Computer-generated display of [g] in *ago*.

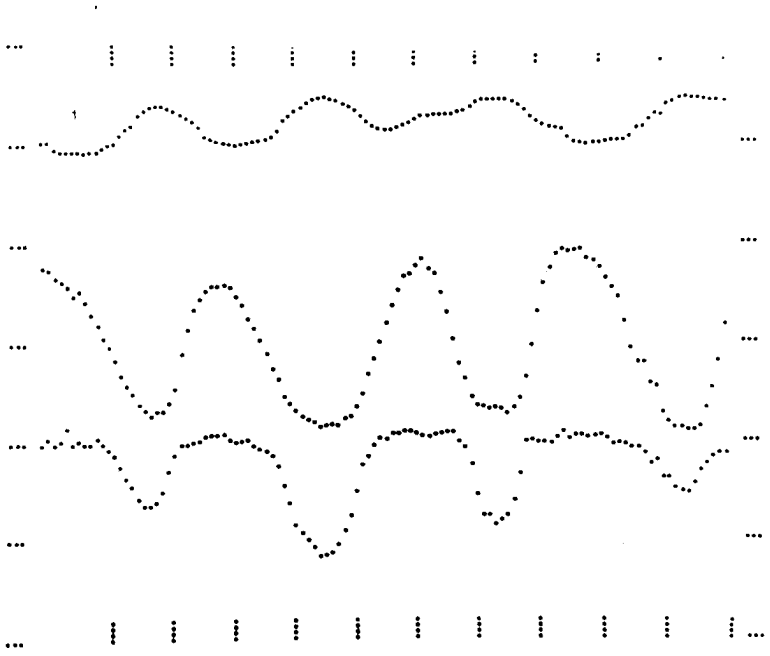


Figure 4. LPC analysis of "We owe you a yoyo".

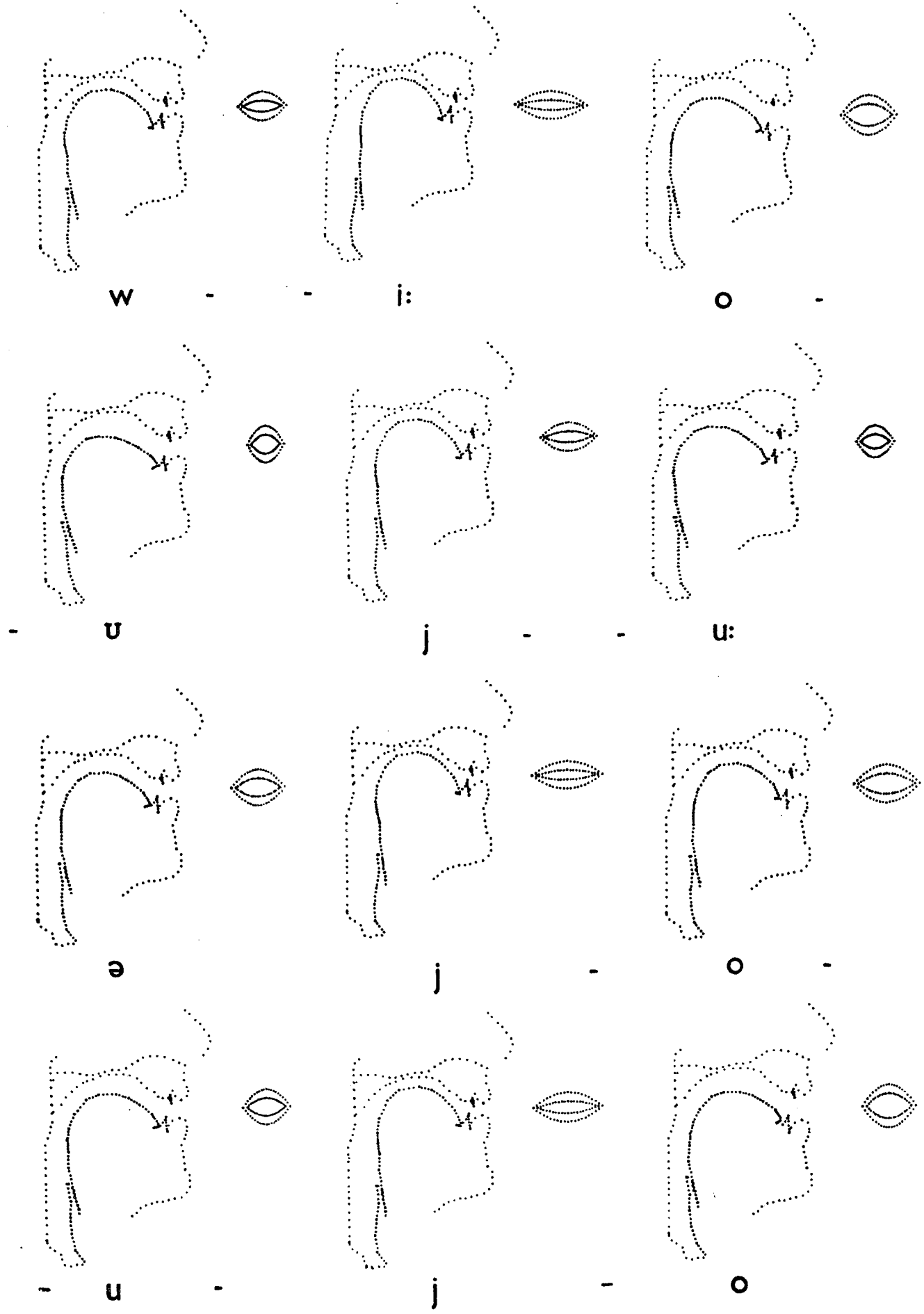


Figure 5. Selected computer generated frames from the utterance "We owe you a yoyo".

The algorithms that were developed for single vowels will work very well for both vowels and glides in continuous speech. Figure 4 shows the LPC formants of the utterance "We owe you a yoyo" A selected corresponding to each phoneme is shown in Figure 5. Note that the system incorporates some kinds of allophonic effects, as for example the different degrees of tongue raising in the initial and medial semivowels in *yoyo*, due to the different degrees of stress. It is also possible to see some coarticulatory effects, such as the anticipation of lip rounding during the first sound in *you*. But some coarticulations are not correctly reflected in our present algorithms; the lip spreading in the second semivowel in *yoyo* is almost certainly inappropriate. However, when we consider the different tongue and lip positions in the first sounds in *we* and in *where* in Figure 6, we see that the tongue position and the width of the lips coarticulate with the following vowel.

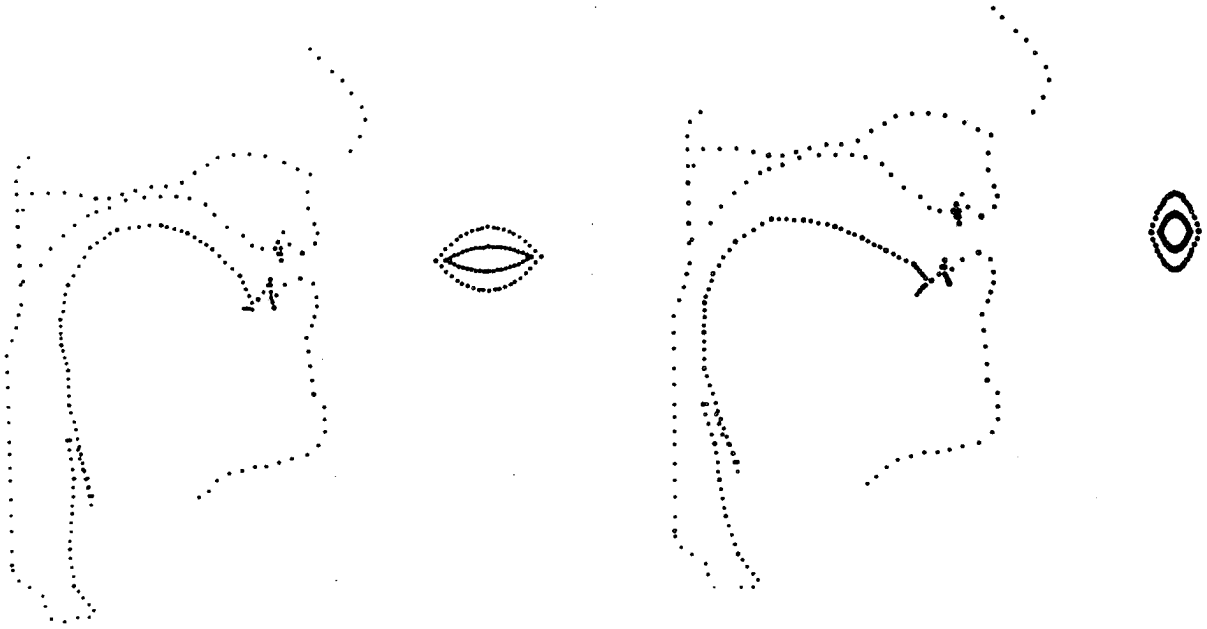


Figure 6. Computer generated [w] in *we* and [w] in *where*.

There was no difficulty in reconstructing plausible vocal tract shapes for each of our three speakers. These may of course not have been the shapes that were actually used, but shapes like those generated could have been used to produce most parts of these utterances.

References

Harshman, R., P. Ladefoged, and L. Goldstein (1977) "Factor analysis of tongue shapes" *JASA*, Vol. 62, 3:693-707.

Ladefoged, P. and R. Harshman (forthcoming) "Formant frequencies and movements of the tongue". In *Frontiers of Speech Communication Research*, ed. by B. Lindblom and S. Öhman.

Ladefoged, P., R. Harshman, L. Goldstein, and L. Rice (1978) "Generating vocal tract shapes from formant frequencies" *JASA*, Vol. 64, 4:1027-1035.

Formant frequencies and movements of the tongue

Peter Ladefoged and Richard Harshman

The most substantial account of articulatory-acoustic relations is clearly that of Fant (1960). Partly because of the excellence of this work, the construction of a model of the vocal tract and the observation of its output has become the predominant method of studying articulatory-acoustic relations. Although many useful observations have been made in this way, it must always be remembered that the validity of the results depends on the validity of the vocal tract model. Furthermore, no one has ever published a study in which they describe a vocal tract model that could generate formants that correspond accurately to the sets of formant frequencies that have been observed in a wide range of vowels spoken by a number of different speakers, using only plausible methods of accounting for the anatomical differences between speakers. There are discrepancies between the observed and the generated formant frequencies even in the case of models that are trying to reproduce the data corresponding to only a single speaker. Thus Mermelstein (1973) had mean errors in F1 of about 10% and in F2 and F3 of about 5%. The work at UCLA (Rice, 1976) showed similar discrepancies between measured and computed frequencies. Because of these problems, it seems advisable to complement vocal tract modeling studies in other ways.

The studies of articulatory-acoustic relations to be reported here use a very different approach. They are essentially studies of the statistical covariation between measured formant frequencies and the corresponding vocal tract shapes as observed through cineradiology. The limiting factors for these studies are the reliability of the data, and the adequacy of the non-linear equations used for capturing the relations involved.

The particular set of articulatory and acoustic data used in these studies has been described by Harshman, Ladefoged and Goldstein (1977), and Ladefoged, Harshman, Goldstein and Rice (1978). A full account of the measurement procedures used has been given in these papers, so that here we need note only that the data consist of measurements of the 10 vowels /i,ɪ,e,ɛ,æ,ɑ,ɔ,o,ɷ,u/ in the words "heed, hid, hayed, head,had, hawed, hoed, hood, who'd" spoken by five speakers. The articulatory data are 18 approximately equally spaced measurements of the sagittal dimension of the vocal tract as seen on the x-ray tracings. The acoustic data are the frequencies of the first three formants.

We will consider two methods of relating these data. First we will examine how the formant frequencies are related to the position of the tongue body taken as a whole. Then we will consider the relation between the acoustic data and particular points on the tongue.

Harshman et al. (1977) have shown that the position of the body of the tongue in these English vowels can be described as the sum of two components, each of which specifies the degree of deviation of the tongue from a neutral, reference, position. The one component specifies the degree of raising or lowering of the front of the tongue (+FR or -FR), and the other component specifies the degree of raising or lowering of the back of the tongue (+BR or -BR). In each case there are concomitant movements of the root of the tongue. In figure 1a the position of the tongue marked +FR has a value of +1.0 of the front raising component (and no back raising component), and that marked -FR has a value of -1.0 of this component. The reference position is shown by the dashed line. In figure 1b +BR and -BR indicate tongue positions with similar positive and negative values of the back raising component, and no front raising component. The tongue positions in a wide variety of vowels can be considered to be the sum of deviations from the reference position specified in this way. There is a high correlation (r = .96) between the observed position of the body of the tongue inthe English data described above and the positions generated by these two components. We can consequently say that these two components account for 92% of the variance in the position of the body of the tongue in this data set.

This finding captures the traditional phonetic belief that there are two primary parameters of tongue shape. These have often been expressed as the height and degree of backness of the highest point of the tongue, without saying exactly how the shape of the tongue as a whole can be calculated from a knowledge of these parameters. Specification in terms of the degree of front raising and the degree of back raising determines not only the highest point of the tongue, but also the position of the whole body of the tongue.

Relating a two dimensional formant space to a two dimensional articulatory space

We may now consider how the observed formant frequencies correlate with these components of tongue position. Ladefoged et al (1978) used a stepwise multiple regression program to determine the relation between various non-linear combinations of the first three formant frequencies on the one hand, and the front raising (FR) and back raising (BR) components on the other hand. The combinations of the first three formant frequencies (F1, F2, F3) that were computed for use as possible independent variables are shown in Table I. Both Fr and Br can be predicted almost equally well from several different subsets of these variables.

The best equations containing only three variables as selected by the stepwise multiple regression between the observed positions of the body of the tongue and the positions reconstructed from the components predicted by these combinations of formant frequencies is .90. Other equations containing a larger number of variables produce only a slight increase in the correlation.

$$FR = 2.309 \frac{F2}{F3} + 2.104 \frac{F1}{F3} + 0.117 \frac{F3}{F1} - 2.466 \qquad (1)$$

$$BR = 1.918 \frac{F1}{F2} - 0.245 \frac{F2}{F1} + 0.188 \frac{F3}{F1} + 9.584 \qquad (2)$$

These equations are difficult to interpret, beyond noting the obvious fact that the tongue positions, assessed across vowels and across speakers, depend on ratios of formants. But it is not immediately apparent why these particular ratios are the most appropriate. (In fact a number of other fairly similar ratios give almost equally good predictions.) The best way of revealing the nature of the relation is by a graphical interpretation.

When one wishes to display graphically the correlation between observed formant frequencies and some aspect of the corresponding articulations, there are problems that do not occur in studies in which a vocal tract model is used to generate a set of formant frequencies. In the latter type of studies the parameters of the vocal tract model are varied in a plausible way, and the corresponding generated formant frequencies are noted. But in our case it is the formant frequencies that have to be varied; and it is not immediately clear how their variation should be constrained. One way is to constrain F3, which has been shown by Broad and Wakita (1977), to be predictable from F1 and F2. The particular equations given by these authors provide only a moderate degree of prediction of F3 in our data (r = .64). For our five speakers F3 may be predicted more accurately (r = .79) from (3).

$$F3 = 0.2191X10^{-3}(F2)^2 - 0.4877F2 + 0.7245X10^{-4}(F1)^2 + 2530. \qquad (3)$$

In order to display the relation between formant frequencies and the corresponding values of the components of tongue shape, a set of formant frequencies was created. F1 and F2 were varied over 50 Hz steps through the range observed in the data, and F3 was determined in accordance with (3). Some impossible combinations of formants were eliminated by requiring the sum of F1 and F2 to be less than 2500 Hz. This set of formant frequencies was then used in euqations (1) and (2) in order to determine the corresponding amounts of front raising and back raising.

The results of these calculations are shown in figure 2, which embeds the formant space for these vowels in an articulatory space defined by the two factors. There appears to be a non-linear but simple

mapping between the two domains. To relate this chart to the four tongue shapes shown in figure 1, solid points indicate the values of the front raising component (+FR and -FR) shown in figure 1a, and of the back raising component (+BR and -BR) shown in figure 1b. Thus, for example, it may be seen that the raised tongue position in figure 1a corresponds to F1 = 550, and F2 = 2000, the calculated F3 being 2450 Hz. The formant frequencies corresponding to each of the other tongue positions in figure 1 can also be seen. In addition figure 2 allows one to calculate a tongue postion corresponding to any permissible set of formant frequencies. Thus for the open circle corresponding to F1 = 410, F2 = 1090, and F3 = 2270 Hz, the tongue shape will have a negative degree of front raising (i.e. some front lowering) and a positive degree of back raising. The tongue positions corresponding to these degrees of front raising and back raisng are shown by the light solid lines in figure 3. The reference position of the tongue is shown by the dashed line, and the combined result of the two components by the heavy solid line. Note that the two components, because they have opposite signs, correspond to opposite deviations in the front of the tongue, and hence tend to cancel one another out in this region. Towards the back of the tongue they add together to make a larger deviation from the reference position.

It should be remembered that equations (1) and (2) are based on the analysis of the observed tongue positions and the corresponding measured formant frequencies in ten English vowels as spoken by 5 subjects. The positions of the lips vary in these vowels, and accordingly the data in figure 2 should be considered to apply only to vowels in which the lip positions varied in accordance with those in English vowels.

The two components of tongue shape are not readily interpretable in terms of the traditional labels 'front' and 'back'. In general, variations in the frequency of the first formant are associated with variations in the back raising component. But it should be noted that positive values of this component occur in so-called front vowels such as [i] and [e], as well as in so-called back vowels such as [u] and [o]. Variations in the frequency of the second formant are associated with both components of tongue shapes particularly when F1 is high. When the first formant frequency is comparatively low F2 is principally associated with variations in the front raising componet. In these circumstances, when the back raising component has a positive value, the front raising component serves primarily to distinguish so-called front vowels from so-called back vowels. In this particular set of data the neutral position of the tongue occurs when F1 = 540, F2 = 1460 and F3 = 2300 Hz. The neutral tongue position is defined here not as part of a uniform vocal tract, but as a reference position which has zero values of both components of tongue shape.

Relating a two dimensional formant space to individual tongue points

We will now consider a different way of associating tongue positions with formant frequencies. Figure 4 shows 16 individual points on the tongue. In order to describe the movements of these points in the English vowels in the data set, sixteen sets of stepwise multiple regresisons were run. In

each case the dependent variable was the distance of one of these points onthe tongue from the roof of the mouth or the back wall of the pharynx. The independent variables were selected by the stepwise procedure from the formant frequencies and combinations of formant frequencies in Table I.

Table II lists the best three-term equations for predicting the positions of the points in each of the ten vowels spoken by each of the five subjects. The correlations between the observed and the predicted position for each point are also shown in this table. These correlations are fairly low in the lower part of the pharynx, where measurement difficulties make the data unreliable. They are also somewhat low in the neighborhood of the velum,where there are only comparatively small movements of the tongue. In addition, the position of the tongue tip is poorly correlated with the formants in these data.

In contrast, certain other positions can be predicted far more accurately. The positions of points 5, 6, 7 on the root and 12, 13, 14 on the front of the tongue can be predicted from the formant frequencies fairly well, at least in this data set. As will be noted below, such equations can probably be applied to other data sets. However the equations are too complex to be readily interpretable. They are best displayed in terms of graphs as in figure 5 and 6, in which the predicted position of the point is shown onthe ordinate,and the variation of F2 is shown on the abscissa, F1 having constant values for any one curve, and F3 being computed as in (3). In these figures the combinations of F1 and F2 have been constrained to the range appropriate for English vowels.

Variations in the position of points 5 and 6 on the root of the tongue (figure 5) are associated with very similar changes in formant frequencies. When the root of the tongue is a comparatively large distance from the back wall of the pharynx (as in [i] or [u]), a considerable range of F2 values may occur. In these circumstances the predicted position of the tongue is more specifically determined by the frequency of the first formant. It is only when the root of the tongue is somewhat closer to the back wall of the pharynx (as in [ɑ]), that F1 and F2 covary. Variations in the position of point 7 (figure 5) are associated with similar movements of the formants, but with covariation being somewhat more extensive.

The movements of points 12, 13, and 14 figure 6) have almost the opposite relation to the formant frequencies. When the constriction is small (as in [i] or [u]) a wide range of F2 values may occur. It is also noteworthy that there may be a considerable change in F1 associated with comparatively small movements of these points on the tongue. Thus, for point 13, if F2 is about 1600 Hz, movements of about 5 mm may be associated with a 200 Hz change in F1.

Some of the combinations of formant frequencies represented in figures 5 and 6 are not exactly those of any vowel in the original data set. Ladefoged et al (1978) have demonstrated that equations (1) and (2) can be used for predicting tongue positions in English vowels other than those in the original data base, as well as in vowels in some other languages. It seems probable that the equations reported here can be similarly extended to other data, but his has not yet been shown. Accordingly, the relations in Figures 5 and 6 should be interpreted with caution when applied to new data, particularly when the data do not correspond to English vowels. But, equally, it should be noted that in this set of data the graphs indicate, very accurately, the correspondence between measured formant frequencies and measured tongue positions, accounting for 88-94% of the variance in the data across five different speakers.

Finally we may compare these two techniques for predicting tongue shapes from formant frequencies. If we judge goodness of prediction simply in terms of the accuracy of the prediction of individual points on the tongue, then obviously a system in which the individual points are predicted by separate equations is more accurate than one which predicts tongue shapes in terms of the front raising and back raising components. But the weakness of the system in which individual points are predicted is that the reconstructed tongue shapes may be irregular. When the underlying components are used in the reconstruction, the tongue shapes are smoother and look more like those that a human being could produce. Both techniques are useful supplements to computer models of the vocal tract, in that they show actually observed correlations between formant frequencies and vocal tract shapes.

Acknowledgements

We are indebted to several members of the UCLA Phonetics Lab who have been working on this project,notably Louis Goldstein and Lloyd Rice. This work was supported in part by USPHS Grant NS09780.

References

Broad, D.J. and Wakita, H. (1977). "Piecewise-planar representation of vowel formant frequencies." *J. Acoust. Soc. Amer.* 62.6, 1467-1473.

Fant, C.G.M. (1960). *Acoustic Theory of Speech Production.* Mouton, The Hague.

Harshman, R., Ladefoged, P. and Goldstein, L. (1977). "Factor analysis of tongue shapes." *J. Acoust. Soc. Amer. 62.3*, 693-707.

Ladefoged, P., Harshman, R., Goldstein, L. and Rice. L. (1978). "Predicting vocal tract shapes from formant frequencies." *J. Acoust. Soc. Amer.* (In press)

Mermelstein, P. (1973). "Articulatory model for the study of speech production." *J. Acoust. Soc. Amer.* 53.4, 1070-1082.

Rice, L. (1976). "A better LASS" *J. Acoust. Soc. Amer.* 60(S1). S78(A).

FL	F1/F2
F2	F1/F3
F3	F2/F3
F1 F2	F1/(F2*F3)
F1 F3	F2/(F1*F3)
F2 F3	F3/(F2*F1)

Table 1. First 12 independent variables used in step-wise multiple regressions. The second 12 are the reciprocals of these terms.

Table II. FORTRAN statements for predicting Y(I), positions of points on the tongue, from formant frequencies, F1, F2, F3. The value of r is the correlation between the predicted position and the observed position across a set of 50 English vowels (5 subjects each saying 10 vowels).

```
S(1) = 1.793 -1.343*F1/F2                                              R=.354

S(2) = 2.857-2.197*F1/F2-354.*F2/(F1*F3)                               R=.474

S(3) = 6.206-3.897*F2/F3-.796*F3/F2-5708.*F1/(F2*F3)                   R=.659

S(4) = 3.2 -3493./F2+427.*F3/(F1*F2)-.000034*F2*F3/F1                  R=.900

S(5) = 3.133-3890./F2+1092000./(F1*F2)-420000./(F1*F3)                 R=.946

S(6) = 2.941-3227./F2+720000./(F1*F2)-.000168*(F1*F3)/F2               R=.956

S(7) = -.092+2.802*F2/F3-1172./F2+206.*F3/(F1*F2)                      R=.944

S(8) = .871+2.584*F1/F3-2107000./(F2*F3)+640.*F2/(F1*F3)               R=.901

S(9) = -3.416+3.34*F1/F2+4.509*F2/F3+.174*F3/F1                        R=.840

S(10) = .39-4071000./(F2*F3)+9524.*F1/(F2*F3)                          R=.726
                +622.*F2/(F1*F3)

S(11) = -.16 +3.724*F1/F2-300000./(F1*F2)+478.*F2/(F1*F3)              R=.827

S(12) = 1.726-1.273*.000001*F1*F2-236.*F3/(F1*F2)                      R=.941
                +.001652*(F1*F3)/F2

S(13) = 8.077 -8.115*F2/F3-1981./F1+2038.*F2/(F1*F3)                   R=.967

S(14) = 7.819 +.763*F2/F1-7.623*F2/F3-.739*F3/F1                       R=.960

S(15) = .044-262000./(F1*F2)+4000000./(F2*F3)                          R=.901
                +.000953*F1*F3/F2

S(16) = .767 +6.026*F1/F2-11.705*F1/F3+.00343*F1*F2/F3                 R=.760
```

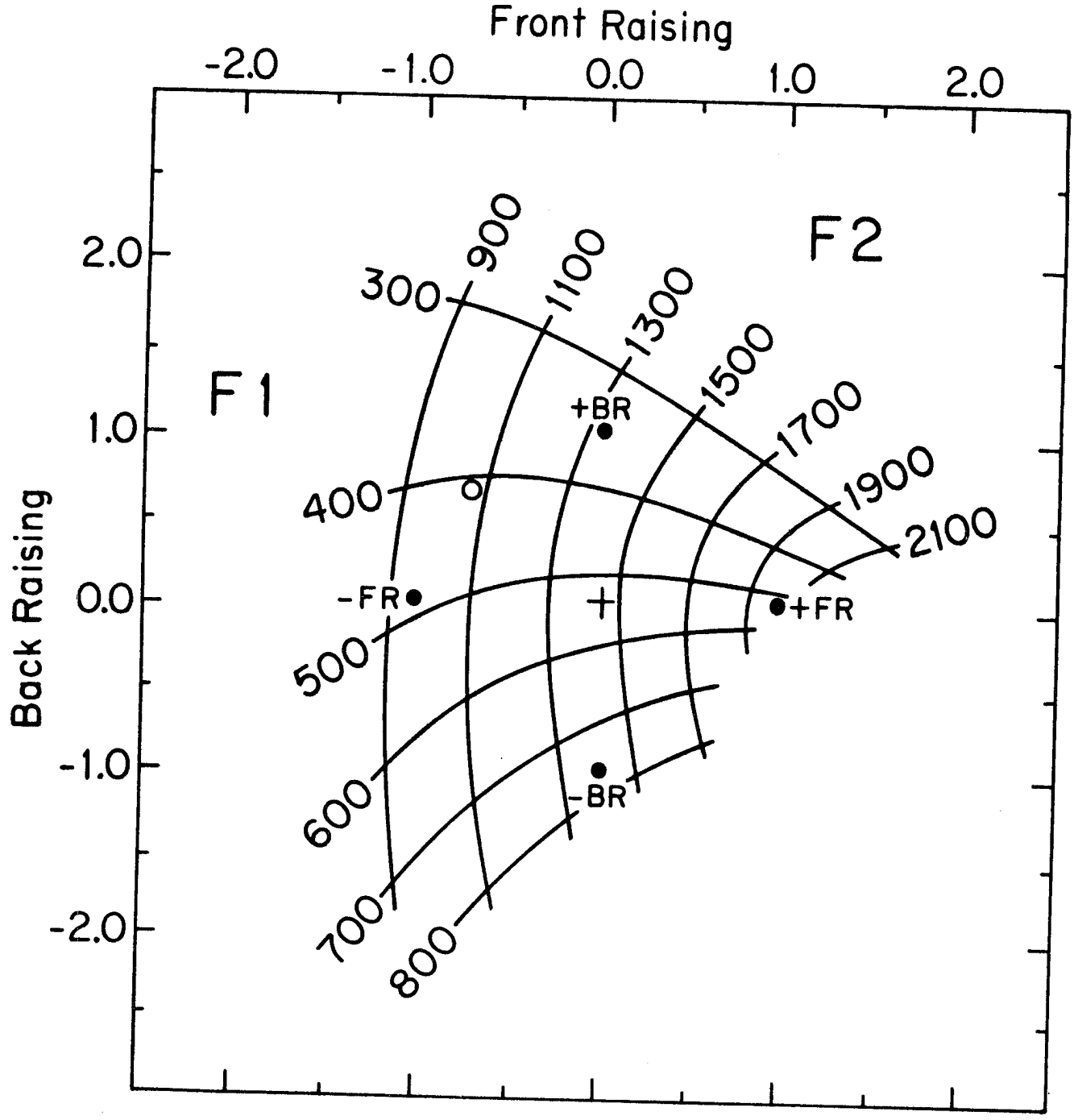


Figure 2. The relation between observed formant frequencies and the front raising and back raising componentsof tongue shape. F3 is calculated from F1 and F2 as shown in (3).

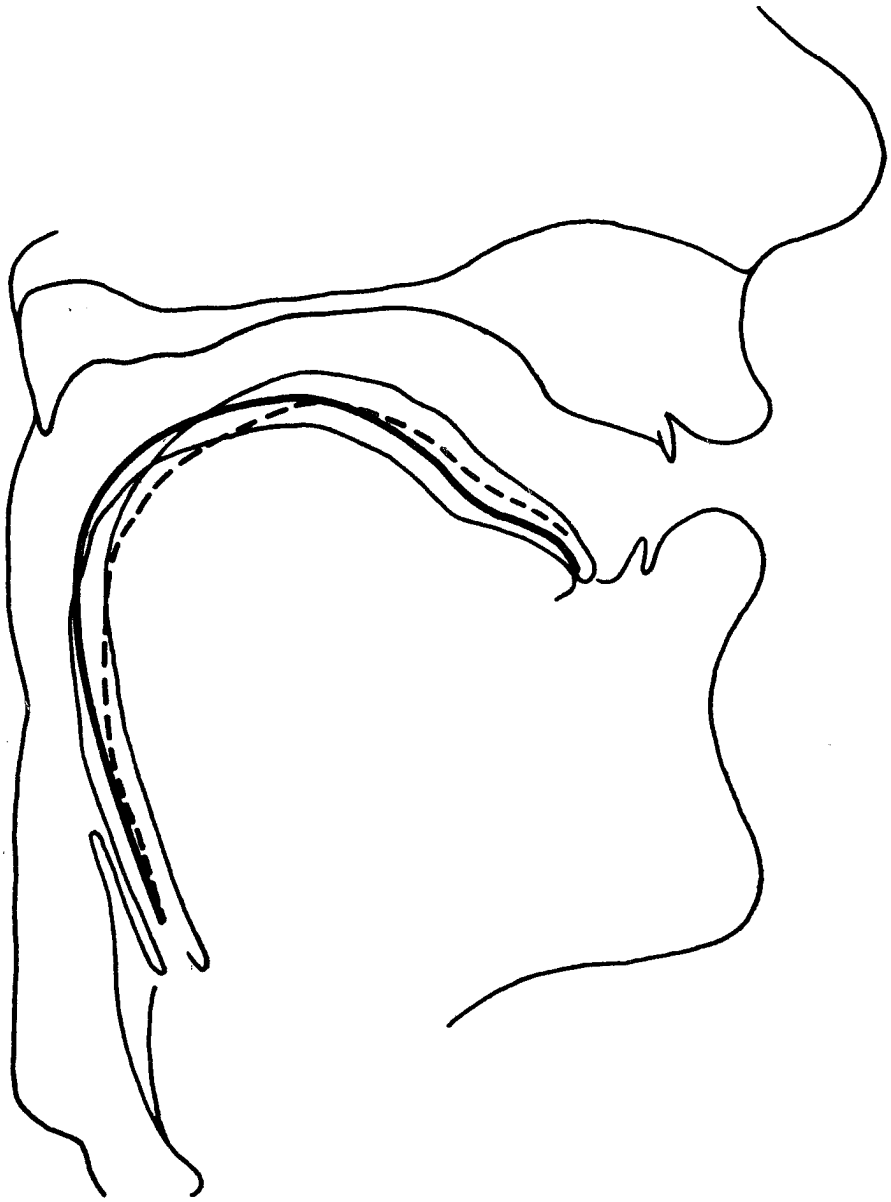


Figure 3. A front raising and back raising component that can be added to form a tongue shape corresponding to F1=410, F2=1090, F3=2270. (Heavy solid line). The dashed line is the neutral, reference, position.

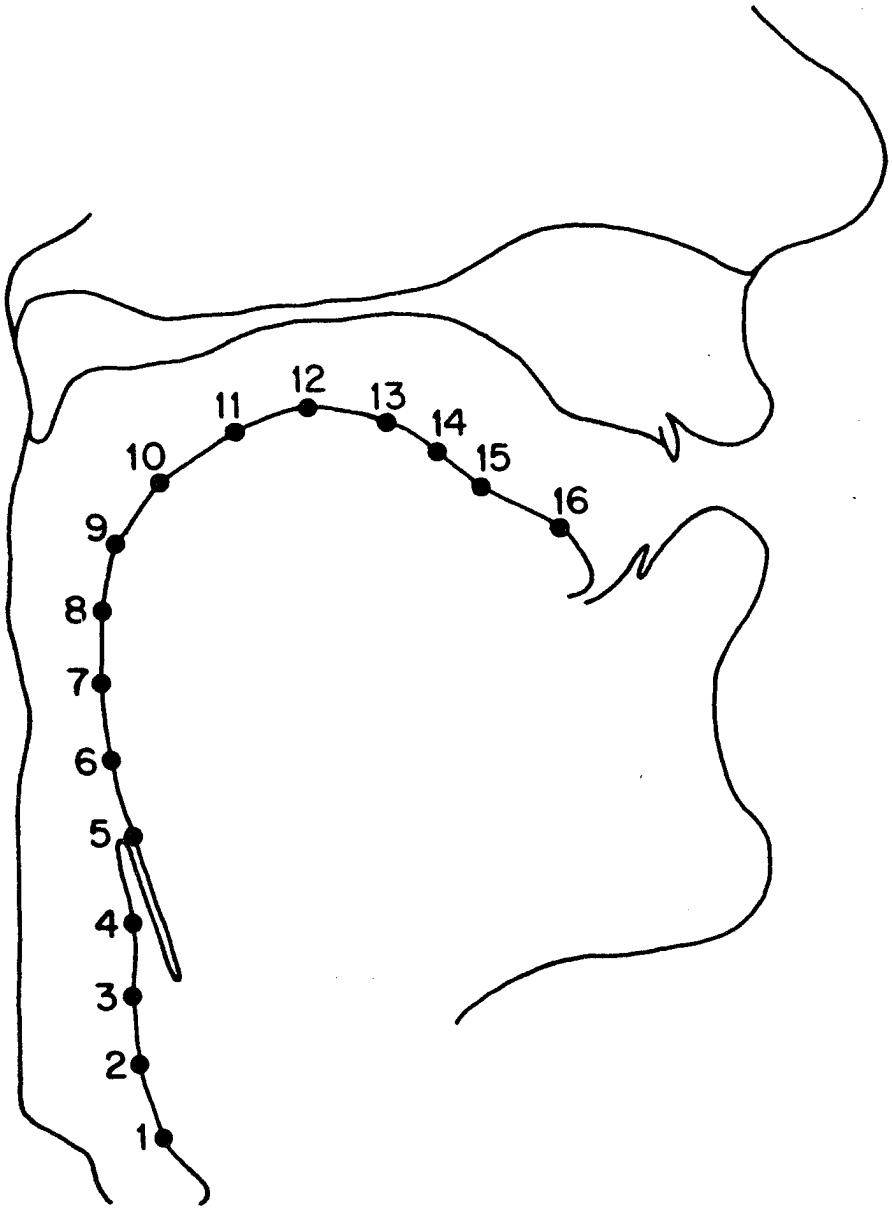


Figure 4. Sixteen points on the surface of the tongue.

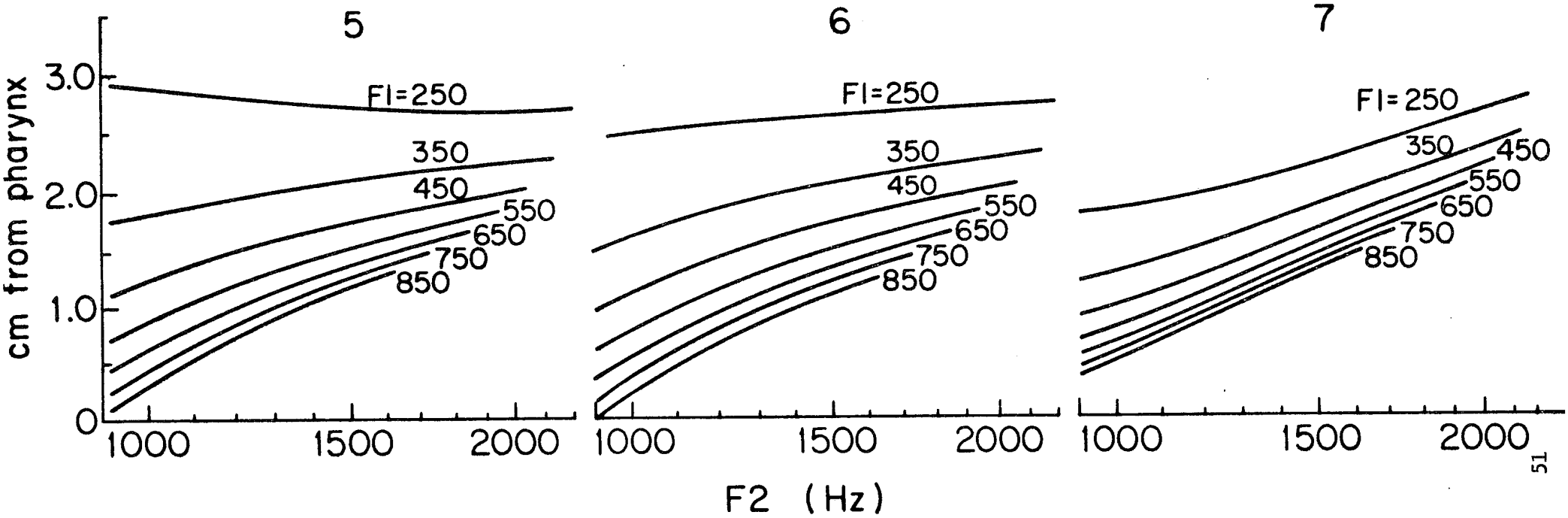


Figure 5. The relation between the positions of points 5, 6, 7 on the tongue and F1 and F2 (F3 being calculated as in (3)).

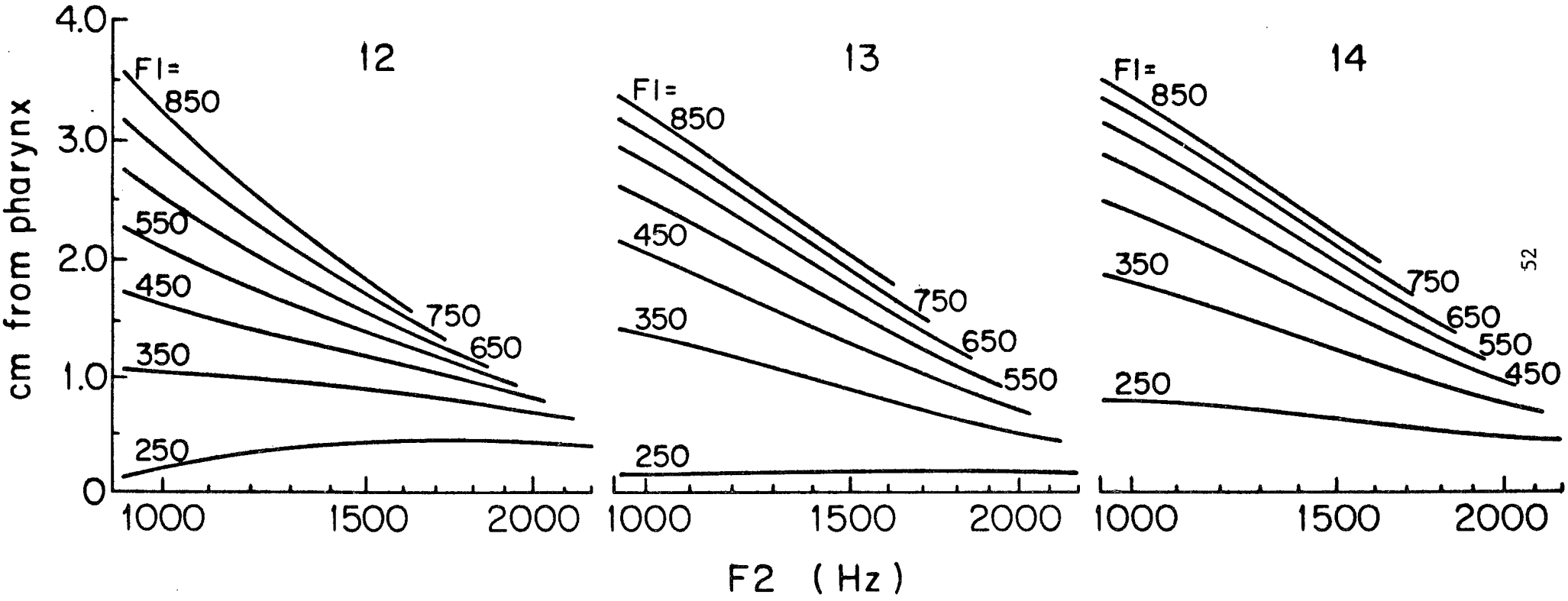


Figure 6. The relation between the positions of points 12, 13, 14 on the tongue and F1 and F2, (F3 being calculated as in (3)).

Where does the vocal tract end?

Peter Ladefoged, Jim Wright and Wendy Linker

[Paper presented at the 95th meeting of the Acoustical Society of America]

The title of this paper is a question that we cannot answer at the moment. So far all we've been able to find out is that our models do not serve as an adequate theory of speech production. We photographed three subjects saying sequences such as [yi, øe, ɯu]. The subjects were all reasonably skilled phoneticians, and they were instructed to produce these sequences by moving nothing but their lips. In a recent paper in the Journal Carol Riordan provided data on speakers of French and Mandarin which contrast vowels such as [i] and [y]. These speakers change not only the position of the lips but also the height of the larynx. Rounded vowels typically have a lowered larynx. We cannot guarantee that our subjects did not alter the larynx position in going from [i] to [y], but at least in the case of subject PL, whose data we will be reporting here, we have his assurance that he could feel no change in the position of anything except his lips. It is quite easy for a trained subject to monitor the steadiness of the larynx position by placing the fingers on the throat. In a future version of this experiment all the subjects will be trained to do this.

While the subjects were being photographed, they were also being recorded. The exact moment corresponding to the photograph could be determined by the click of the shutter on the recording. Acoustic analyses were made of these moments, and the formant frequencies found by an LPC technique. The area of the lip opening was determined from tracings of the photographs.

Let us first consider what standard acoustic theory would predict. The first figure is the well known diagram by Fant showing the variation in the formants for five different lip positions for various different tract shapes. As you can see, if the minimum area of the vocal tract is about 14 cm from the glottis, as it would be for [i], variations in lip rounding have the greatest effect on the third formant. There is very little change in either the first or the second formant as lip rounding increases. This figure is a little hard to read for our present purposes, but it provides a useful reminder of the general situation. Now let us look at Figure 2 which shows just the calculations that are relevant in our situation. The points on the left of the figure indicate the frequencies of the first three formants when the subjects said [y]. We do not know the tongue position that the subject used. But we know from the photographs that the lip area was a little less than 0.15 sq cm. We also know, from

previous x-rays and other measurements that this subject's vocal tract is about 19 cm from the glottis to the upper teeth; and in the vowel [y], which has a very small lip opening, the end of the vocal tract is located one cm beyond the upper teeth, so that the total length is 20 cm. Using these values on an interactive computer program we determined a plausible vocal tract tract shape that could have produced the observed formant frequencies. Then using an algorithm published by Fant and Liljenkrants we calculated what the formants would be given the same vocal tract shape but with the variations in lip opening observed in the experiment. These are the other points on the figure. As you can see, increasing the lip area has very little effect on F1 or F2, but causes a considerable increase in F3.

Now having seen what our model - essentially Fant's model - predicts should be the effect of changing the lip area, let us see what the observed formant frequencies actually were. Figure 3 shows not only the calculated formant frequencies just described (the solid circles), but also the actually observed formant frequencies. As you can see, the square points representing the observed formant frequencies lie right on top of the calculated formant frequencies for the first formant. But they are way off for the second formant and not exactly right for the third. Lip rounding has a considerably greater effect on the second formant than would be predicted for this vowel.

The same kind of difference between observation and prediction occurs for other vowels. Figure 4 shows the situation for the sequence [ø e]. Again the solid circles represent the calculated values, and the squares show the observed formant frequencies. And again the second formant rises more rapidly than is predicted.

Our first attempt to explain this difference was in terms of the length of the vocal tract. Figure 5 shows the area function that was used for the [y i] sequence. The solid line indicates the areas of the 18 section tube used for [y], with a very small opening at the lips: the solid line followed by the dotted line shows the area function assumed for [i] in the previous calculation. It is a simple increase in the area of the final section, the whole of the rest of the vocal tract being assumed to be the same. We thought that this vocal tract shape had given us the wrong values of F2 and F3 shown in the previous diagram because we had not shortened the vocal tract appropriately. This shortening must, of course, be done only at the front end, because it is only the lips that have a different position. Accordingly we tried shortening the vocal tract by almost 2 cm, deleting one section so that we had 17 tubes instead of 18, and making the final seventeenth section with the observed lip area but very close to the penultimate sixteenth section, as indicated by the line of asterisks. But it turns out that if we calculate the formants for a shortened vocal tract with an area function as indicated by the solid line and the asterisks, there is only a very small change - less than 5 Hz - in the values of any of the formants.

Obviously our models are wrong, perhaps, because we - following Fant - have a wrong model of the impedance at the lips. This is obviously the next thing we should try changing. But perhaps it is also because we should not think simply of plane waves moving down the vocal tract. When the lips are spread, as in [i], the vocal tract is open to the air at a point considerably behind the upper front teeth, and there may well be transverse waves to be taken into account. Something of this sort seems indicated by the data in Figure 6, which shows a spectrogram of this subject saying [iy]. Notice the great discontinuities in the third and fourth formants. Obviously adding lip rounding does not cause just a steady lowering of the frequencies of the higher formants. Try and consider yourself as a formant tracker, following along the third and fourth formants, and you'll find you have great difficulties. Life is not as simple as the Acoustic Theory of Speech Production would lead us to believe. Any suggestions on how it really is will be gratefully received.

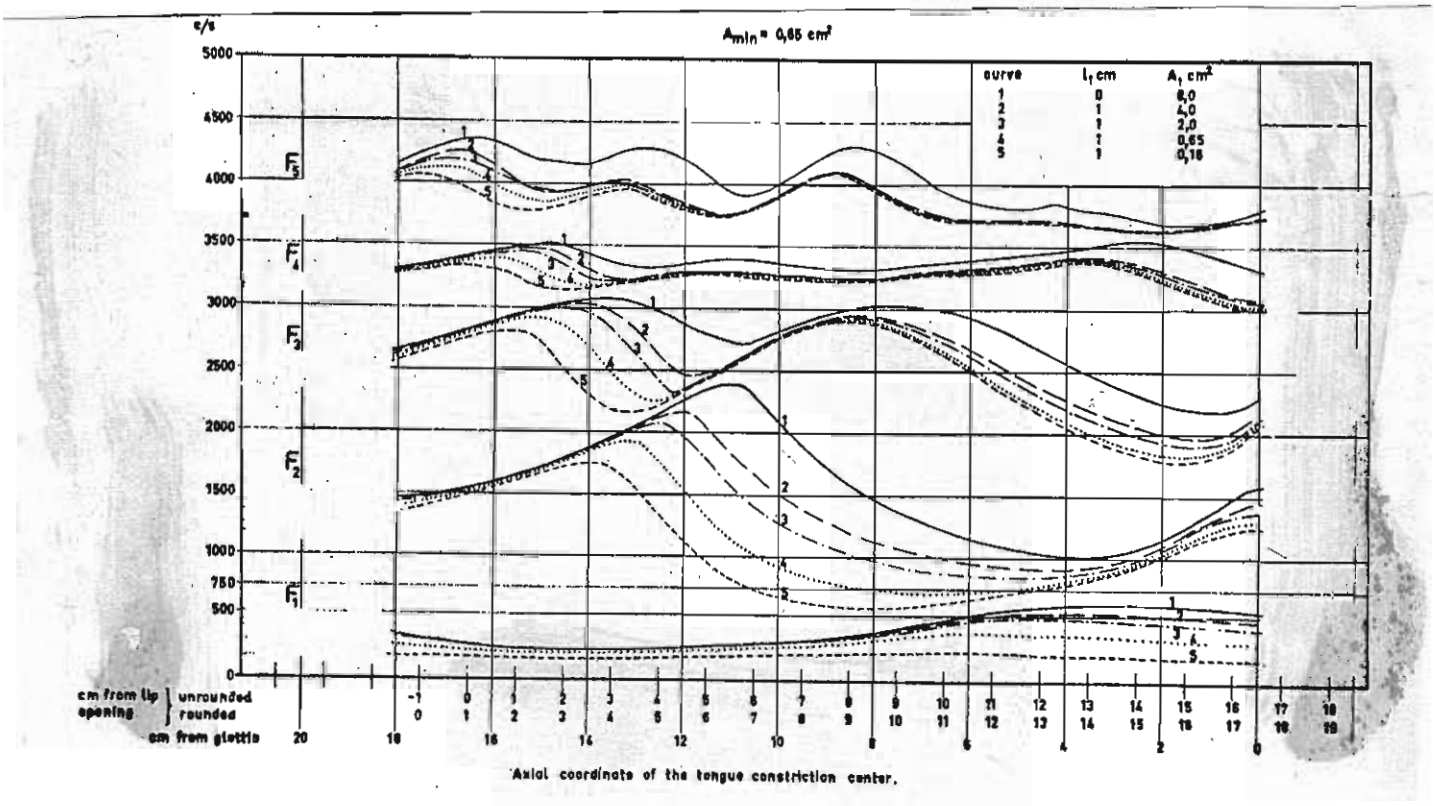


Figure 1. Nomograms of the first five resonance frequencies of the model. Tongue constriction area A_{min} = 0.65 cm^2. Solid curves denoted 1 pertain to no lip section. Curves 2-5 represent an addition of a lip section with increasing degrees of rounding, i.e. decreasing area of the lip section (Fant, 1960).

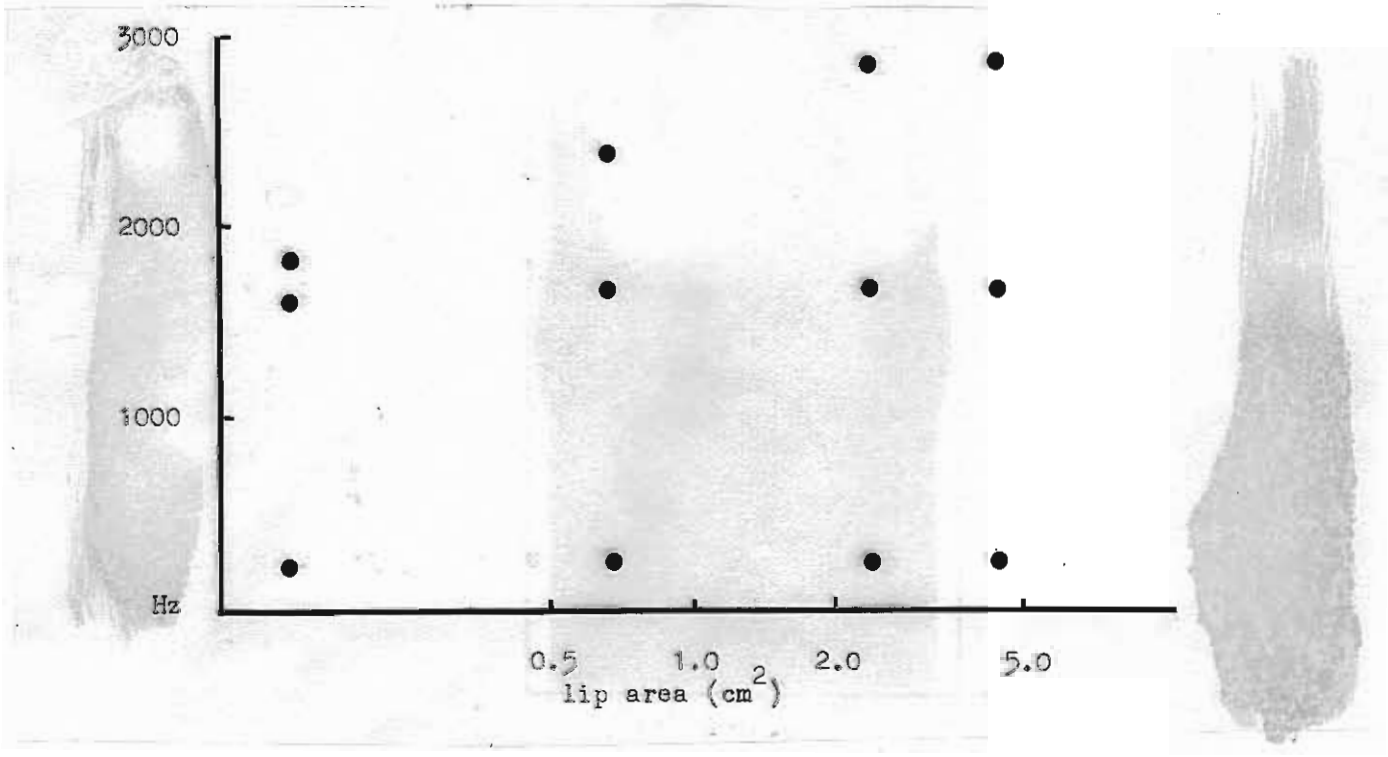


Figure 2. Lip spreading effect on the first three formants in the sequence [yi] as predicted by standard acoustic theory.

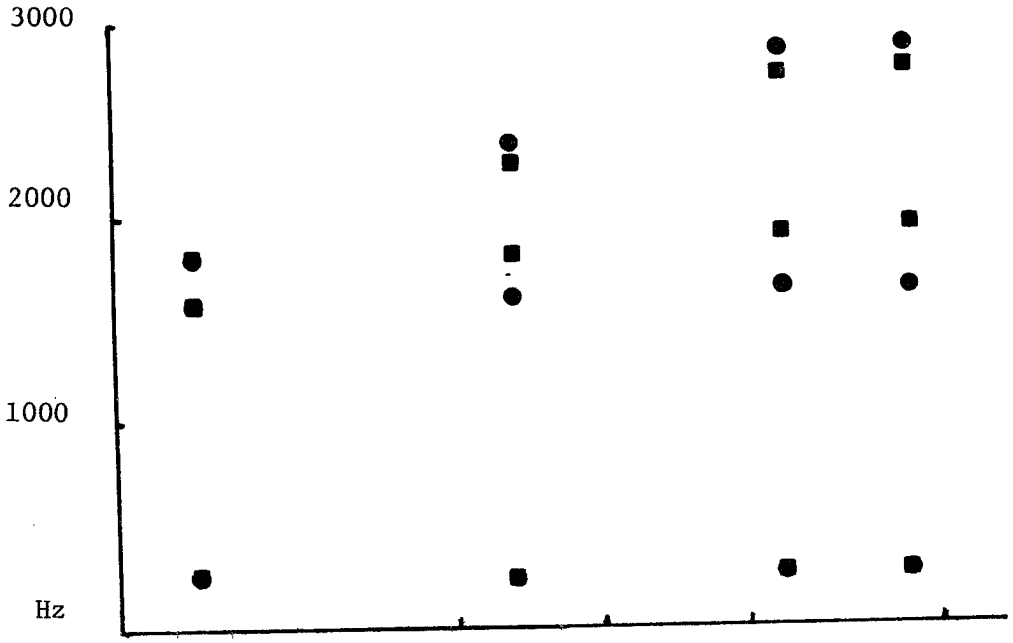


Figure 3. Comparison of actually observed effect (squares) and predicted (circles) effect of lip spreading on the first three formants during the sequence [yi].

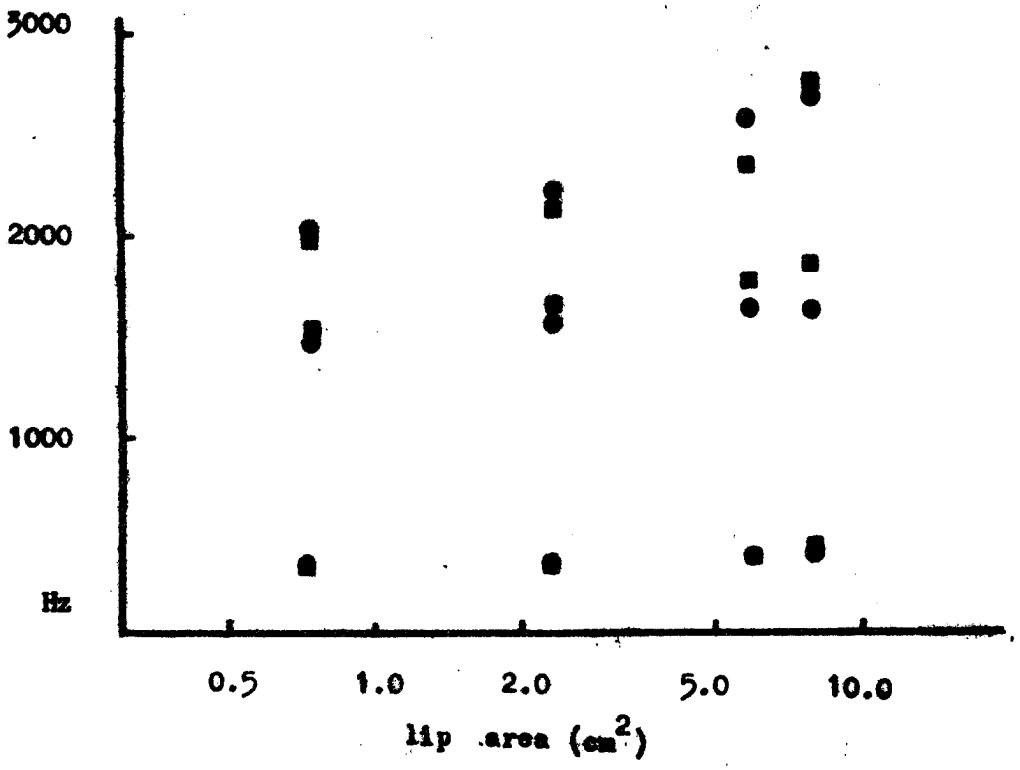


Figure 4. Comparison of actually observed effect (squares) and predicted effect (circles) of lip spreading during the sequence [øe].

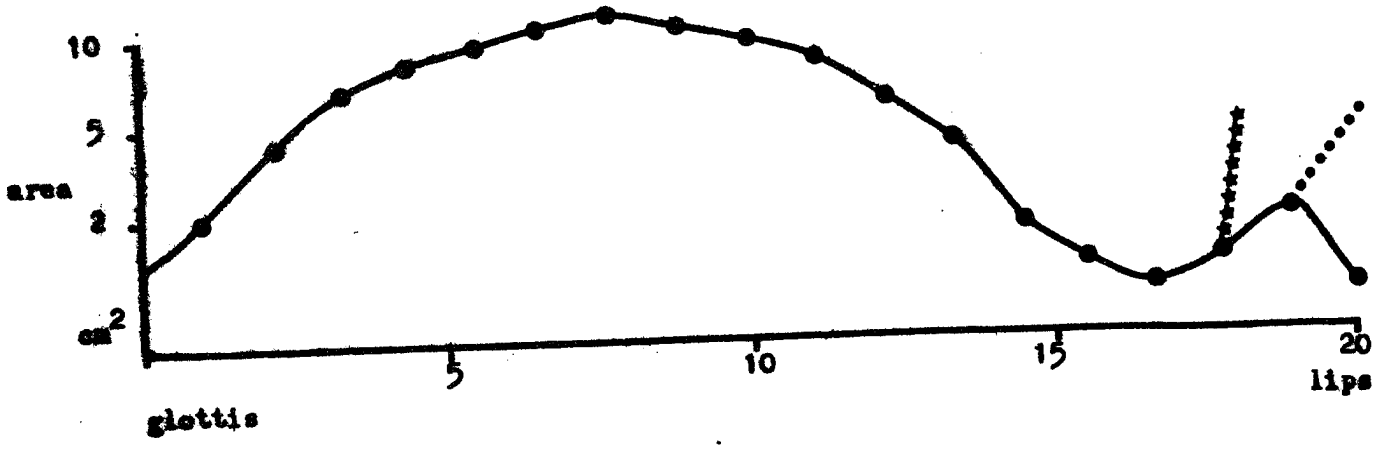


Figure 5. Area functions for a high front vowel with varying lip openings (solid vs. dotted lines) and varying vocal tract lengths (asterisk vs. dotted lines).

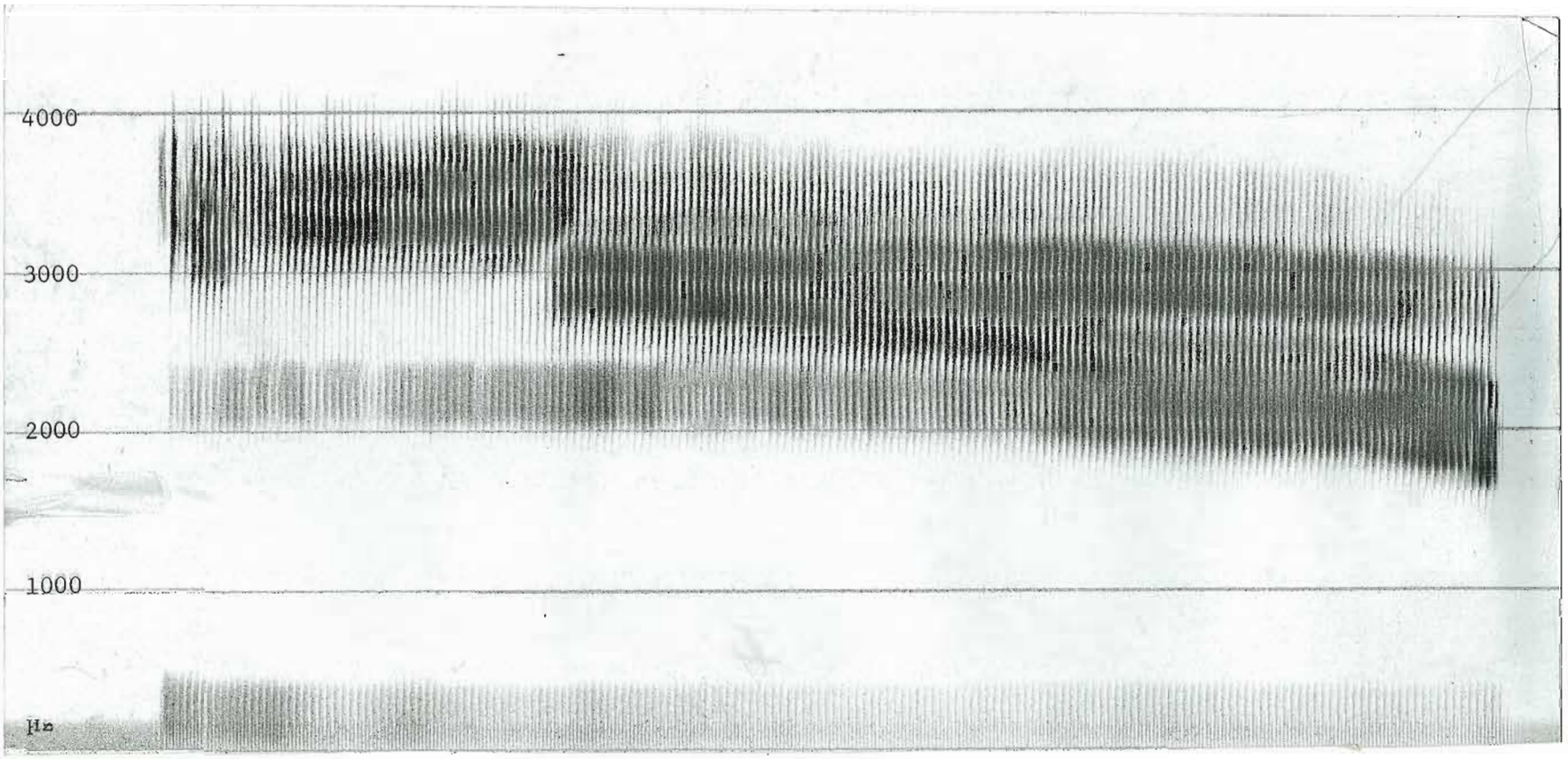


Figure 6. Spectrogram of the sequence [iy].

*The Epiglottis as an Articulator**

Asher Laufer and I. D. Condax

Introduction: Survey of the Literature

We believe that the importance of the epiglottis in speech has been generally underestimated in the phonetic literature. Our evidence leads us to conclude that the epiglottis is an active articulator in the production of pharyngeal consonants and that it is involved in the production of the vowel [a], and in the production of whisper.

To the extent that the epiglottis is mentioned at all, it is generally said to have no speech function. Hardcastle (1976) mentions the function of the epiglottis only in the following terms: "The primary function of the epiglottis is to close off the entrance to the larynx during swallowing, to prevent food passing into the trachea (p. 60)." We find no other function attributed to the epiglottis in his text.

The production of pharyngeals (which according to our observations, as we shall show, involves the epiglottis and not the tongue root or tongue dorsum) is attributed by several phoneticians to an articulation between the tongue and the pharynx. Heffner (1964), for example, says: "By drawing the body of the tongue back toward the posterior wall of the pharynx with very considerable force, one can produce a constriction of the pharynx slightly below and behind the extreme edge of the velum. No stop consonant is produced by such a constriction, but distinctive pharyngeal fricatives are thus produced (p. 152)."

Chapman (1973) writes: "The epiglottis does not seem to have any function in speech (p. 6)." In the description of pharyngeals, he writes: "The tongue root is drawn backward towards the pharyngeal wall to cause a constriction of the pharyngeal cavity. This has the effect of flattening the tongue against the bottom of the mouth and pushing the tip forwards (pp. 13-14)." The sagittal section of a pharyngeal fricative he presents shows the tongue dorsum near the pharynx, the epiglottis far from it and unrealistically small and low in the pharynx. The epiglottis is also much farther from the wall of the pharynx than is the tongue in his illustration. (Chapman (1973) p. 29, Diagram 13).

O'Connor (1973), too, says: "The very back of the tongue may, as we have seen, be pulled backward into the pharynx, thus modifying the latter's

* Asher Laufer is at the Hebrew Language Department, Hebrew University, Jerusalem, Israel. I. D. Condax is at the Department of Linguistics, University of Hawaii, Honolulu, Hawaii 96822.

shape and affecting sound quality: this happens in some pronunciations of ah. Pulled farther back still, the tongue may come so close to the back wall of the pharynx that air passing through causes friction (Figure 13); two such sounds, one voiced, one voiceless, occur in Arabic. Sounds made in this way are known as pharyngeal sounds; strictly speaking they should be referred to as linguo-pharyngeal, but in all cases where the tongue is involved we generally omit 'linguo-' unless it is specially necessary (p. 42)." O'Connor's diagram of pharyngeal consonant articulation showing that in his opinion the consonant constriction is formed between the back of the tongue and the pharyngeal wall.

Catford (1977) distinguishes two types of pharyngeal articulation. One he terms faucal, in which "the part of the pharynx immediately behind the mouth is laterally compressed, so that the faucal pillars move towards each other. At the same time the larynx may be somewhat raised. This appears to be the most common articulation of the pharyngeal approximants [ħ] and [ʕ]." The second type of pharyngeal is said to be where "the root of the tongue, carrying with it the epiglottis, moves backwards to narrow the pharynx in a front-back dimension" (1970, 163). These definitions apply to pharyngealized sounds as well: "Pharyngealized sounds involve some degree of contraction of the pharynx either by a retraction of the root of the tongue, or by lateral compression of the faucal pillars and some raising of the larynx, or a combination of these" (1977, 193). Although he mentions the epiglottis, he considers its actions to be entirely dependent on the movements of the tongue, and accords the epiglottis no status as an independent articulator.

Our observation is that, among the other manners of articulation in the pharynx (fricatives and glides), pharyngeal stops are also found in careful pronunciation in Hebrew; Malmberg (1963) recognizes the possibility of making pharyngeal stops (without saying what language might have them): "A stop may also be realized in the pharynx (pharyngeal stop)" (p. 43). This opinion is shared by Ladefoged (1971) to a large degree, as discussed below. Others have said that a pharyngeal stop is impossible: Brosnahan and Malmberg (1970) speak of pharyngeals as "radico-pharyngeal, in which the root of the tongue articulates toward the rear wall of the pharynx" (p. 45), with the following limitation: "The complete blockage of the supraglottal tract which is characteristic of a plosive can be formed at all the positions of articulation detailed above save between the radix of the tongue and the rear wall of the pharynx" (p. 107).

According to Ladefoged: "In the pharyngeal area, however, no language uses stops (most people cannot make them). . . . Even fricatives are not very common" (p. 41). His diagram of the places of articulation shows the back of the tongue as the articulator that approaches the pharyngeal wall for the production of pharyngeals (p. 36).

In the same publication (1971,62), however, Ladefoged comes curiously close to seeing the vowel [ɑ] as a pharyngeal. In the course of giving symbols for secondary articulations (Table 39) he gives the symbol superscript [ˤ] as the symbol for pharyngealization. He describes this sound as involving "retracting of the root of the tongue", however, not retracting the epiglottis.

Hockett (1958) speaks of pharyngeal stops and voiced and voiceless pharyngeal fricatives for Arabic, but attributes these to an articulation involving the tongue and the pharynx (p. 66).

In this introduction, we have quoted only a sample of the statements on the epiglottis in the current literature of phonetics. Neither author has in any case found any previously expressed opinion that the epiglottis is involved in the production of pharyngeal consonants, and only one reference to the epiglottis in the production of a vowel (Russell (1931), to be discussed later in this paper).

We get the impression from our survey of general phonetic literature that there are widespread misunderstandings about the role of the epiglottis in speech, and especially about the nature of pharyngeal articulation. There is a difference of opinion on a related question of possible degrees of closure, some phoneticians saying full closure is impossible, others considering it to be possible or even in actual use. To anticipate our conclusions, our observations lead us to conclude that the epiglottis is an active articulator involved in producing [ɑ], pharyngeals, which are produced as glides, fricatives and stops, and that the epiglottis is involved in whisper.

Equipment, methods, subjects and texts

Color videotape and still films, both with audio, were made using a fiberscope focused on the epiglottis and adjacent structures (tongue root, pharynx, and larynx). The fiberscope consisted of a light source, a fiberoptic bundle to transmit the light, and another group of fibers with a lens on each end to transmit an image from the tip of the bundle back to a video camera so that it could be displayed on a small screen. The image could also be recorded on a video tape recorder. The end of the fiberscope, after being introduced through a nostril and the nasal passage into the upper pharynx, was positioned slightly below the velum and above the surface of the tongue. From this position, the posterior wall of the pharynx, the lateral walls of the pharynx, the aryepiglottic folds, the vocal folds, the upper edge of the epiglottis, and occasionally the posterior portion of the tongue (in articulations with a particularly retracted tongue position) were clearly visible.

On the screen the epiglottis is seen at the bottom, the arytenoid cartilages and posterior pharyngeal wall at the top. The pictures on the video screen occupied about one fourth of the total available video screen area. This image was used for both monitoring during recording

and, with appropriate stop frame and slow motion, for data analysis. Illustrations for this paper were made by photographing the video screen using a 35mm still camera: see the appendix for a list of equipment. The conclusions from fiberscopic observations are supplemented by the dissection of the musculature and cartilages of the epiglottis and larynx in a cadaver.

The subjects were nine native speakers of Hebrew. With one exception, they all spoke the eastern dialect natively, and all of them pronounced the Hebrew pharyngeals of that dialect well. One of the subjects, a native speaker of the western dialect, but since childhood able to speak the eastern dialect, pronounced the pharyngeals in a way that by fiberscopic examination, ear, and acoustic analysis resembled the pronunciation of the native speakers of the eastern dialect. The subjects, and/or their families, came from Morocco, Egypt, Syria, Iraq, Kurdistan, Yemen and Israel.

The subjects were asked to read (or recite, if the lighting and other factors in the physical set-up of the experiment did not at times permit them to read) nonsense syllables, words in a list, and the story of the North Wind and the Sun.

Four subjects spoke the words in a frame sentence. The prepared material was in each case supplemented by free conversation, with the equipment in place. Thus there is a considerable amount of material, ranging from slow and careful to rapid and casual. A total of 100 minutes or recording (all subjects combined) was obtained.

The copies of the videotapes on which this work is based and an edited version presenting illustrative examples with explanatory titles are available to interested colleagues for viewing at the Hebrew Language Department of the Hebrew University in Jerusalem, at the Department of Linguistics. Phonetics Laboratory at the University of Hawaii,and at the Department of Linguistics Phonetics Laboratory at UCLA.

We performed all analysis based on the tapes with coordinated audio, using normal speed, slow motion, and stop-frame viewing. The degradation of quality in the still photographs reproduced here (black and white photograph of color video, and offset reproduction of the photograph) does not reflect the quality of the original, even in stop frame, which itself contains only half the information in terms of video scan of the moving picture. The presence of color and motion, in addition, contribute a great deal to the clarity and identifiability of structures in the total picture. The outlines accompanying the photographs are tracings made from the photographs printed here, but the exact location of the lines in the less clear areas were decided on using all available information, including repeated viewing of the word containing the example in slow motion and frame-by-frame display. A key to the interpretation of these tracings is given in figure 1.

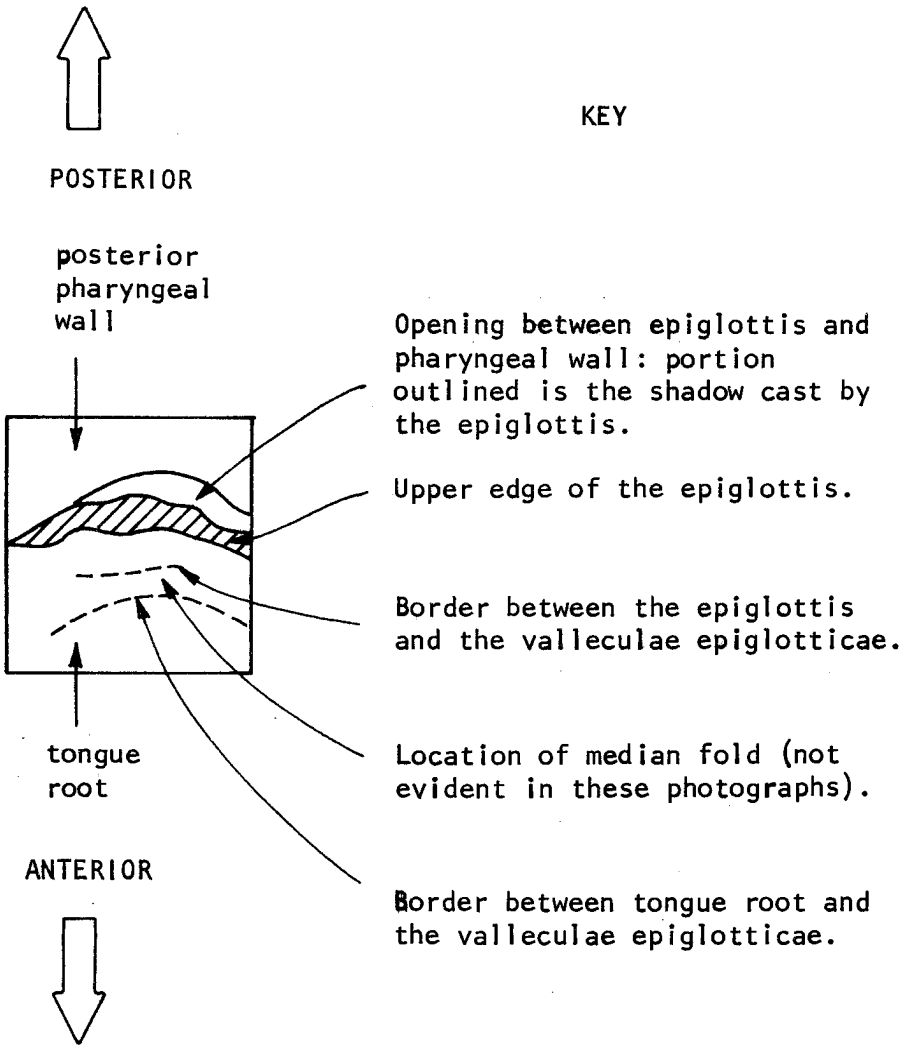


Figure 1. Key to the interpretation of tracings of the photographs of the epiglottis.

Results

We observe the following movements of the epiglottis:

(1) In pharyngeals, we observe that the epiglottis moves back independently of the tongue root (which usually is so far forward as to not show in our picture). The epiglottis either forms a narrow opening, close to the posterior wall of the pharynx and touching the lateral walls of the pharynx, or it forms a complete closure against the pharynx. We examined the epiglottis in our own detailed dissection of one cadaver and we also made briefer examinations of approximately 35 other cadavers dissected by medical and dental students. In the dissected cadaver material, we observed that when an epiglottis reaches the posterior wall of the pharynx, as it does in the production of pharyngeals in vivo, it also presses against the apex of the arytenoids; and that when a narrow opening is formed, the epiglottis presses less strongly against the apex of the arytenoids. The musculature in a position to pull the epiglottis down and back as seen in the cadaver material is more substantial than it appears in illustrations in anatomical atlases we have examined;in particular the aryepiglotticus is continuous with the thyroepiglotticus and thyroarytenoideus,and fibers of all three can contribute to pulling the epiglottis down and positioning the arytenoids under it. We found no evidence for glossoepiglotticus or for pharyngoepiglotticus in the detailed dissection (in this cadaver atrophic changes of age are expected), that is, no muscles opposed the action of the muscles pulling the epiglottis down and back. A less detailed dissection of 3 additional cadavers also failed to yield evidence for these muscles.

(2) In the production of the vowel /ɑ/ we observed that the epiglottis moves substantially farther back than it does for the production of any other vowel except for isolated occurrences of /o/. In these vowels, the side edges of the epiglottis touch the lateral wall of the pharynx, so the narrowest point in the opening is formed by the pharynx and the epiglottis: the tongue is not directly involved in making the maximum stricture.

(3) During the production of whisper we observed that the epiglottis was generally farther back than during normal voice, but in this more retracted position, it continued to make the same kinds of motions it did during normally voiced speech.

Discussion

(1) *Pharyngeals*

We believe the epiglottis is an articulator in the production of pharyngeal consonants for two reasons: the first is that the epiglottis assumes positions near the pharyngeal wall that we believe have acoustic consequences. The epiglottis is capable of assuming a range of positions from completely against the pharyngeal wall (complete closure) through a relatively forward position near the tongue (complete opening). In other

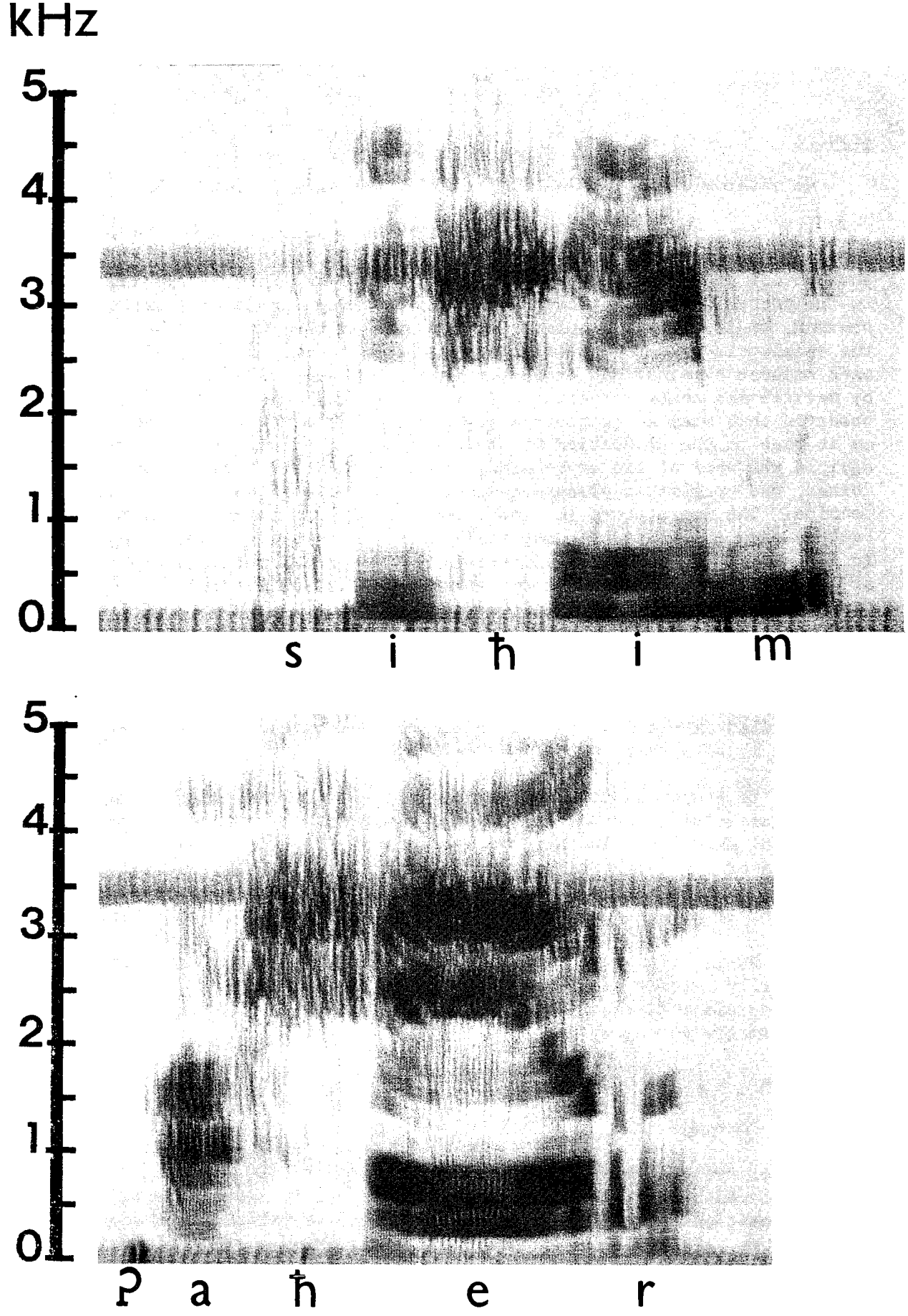


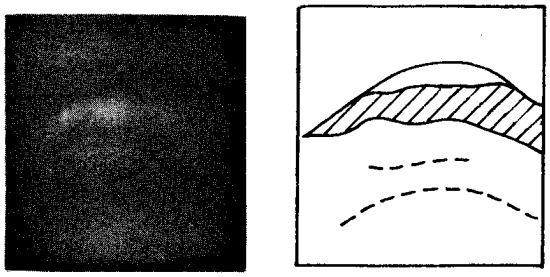
Figure 2. Spectrograms showing the voiceless fricative /ħ/. (The horizontal line at approximately 3.5 KHz is caused by noise from the fiber-optic light source.)

words it forms a major part of an acoustically significant *constriction*, at its minimum extent affecting vowel quality, a tighter constriction producing *friction*, and when maximally constricted forming a *closure* against the pharyngeal wall. In our spectrograms we observe that the transitions (at the beginnings or ends of the vowels adjacent to the pharyngeals) show that some articulator must be moving, and because the transitions are typical of pharyngeals in other published data, we know that the movement must be in the pharynx. When using the fiberscope, we look at the pharynx during the production of these sounds; the only structure we see moving to any large extent is the epiglottis. Smaller movements of the lateral pharyngeal wall, bringing it against the lateral edges of the epiglottis, are also seen. The back of the tongue is seldom seen, and when it is visible in our field of view, it is nevertheless much farther from the pharyngeal wall than is the epiglottis. Thus "pharyngeal" articulations are clearly made by the epiglottis: in all cases we see it moving towards the posterior pharyngeal wall.

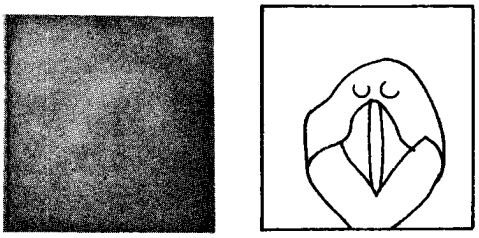
The second reason for believing that the epiglottis is an articulator is the fact that it moves independently of the tongue to a large degree. We want to emphasize that the epiglottis is not just pushed back by the tongue, for, if it were, one might possibly consider the tongue the articulator and the epiglottis merely an insignificant, because accidentally intervening structure. On the contrary, the tongue root is generally some distance from the epiglottis during the production of pharyngeals and it is usually outside the field of view of the fiberscope. The epiglottis folds down and back, away from the tongue. In a number of cases the epiglottis moves back, then a short while later the tongue root moves back after it. The movements are clearly not made simultaneously.

As mentioned above,study of some 36 cadavers shows that contraction of the aryepiglotticus and thyroepiglotticus is probably adequate to account for this folding. These muscles are of adequate size and have suitable angles of pull.

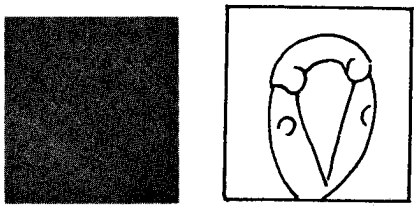
The voiceless pharyngeal fricative /ħ/ is articulated by the epiglottis. In all cases /ħ/ is a fricative, as can be readily heard, and as can be seen in our spectrograms, two of which are shown in figure 2. In some subjects, such as that in figure 3(a), the opening between the epiglottis and the posterior pharyngeal wall is narrow enough that this constriction is obviously the source of the friction In other subjects, however, the opening is relatively large as in figure 4 and we consider it unlikely that this constriction accounts for the observed friction. In these cases, we suppose that the friction results from air passing through the constriction between the base of the epiglottis and the arytenoids. Possibly the fat pad at the base of the epiglottis just above the level of the top of the arytenoid cartilages is involved in creating the constriction. The arytenoids in one subject (Z) are seen to move forward meeting the epiglottis as the vocal tract moves to make the constriction. Such shifting and lifting of the larynx and approximation of the apex of the arytenoids to the base of the epiglottis has been observed in a lateral-view cineradiographic study of another language (Traill (1978)). In any case it is clear that the epi-



(a) [ħ] in [koħo]

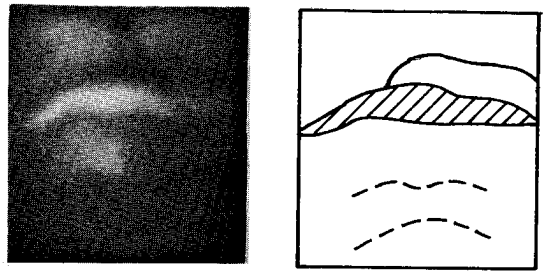


(b) exhaling between
[raʔa] and [raʕa]



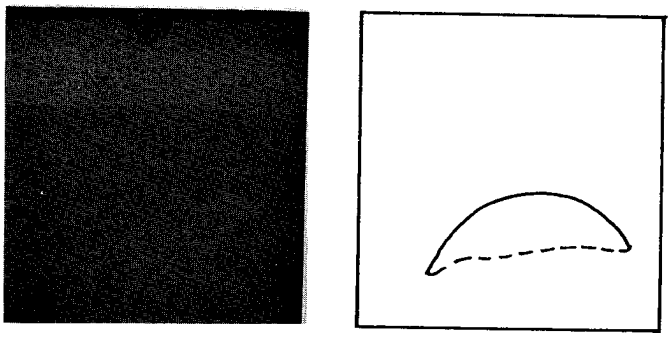
(c) inhaling between
[raʕa] and [raʔa]

Figure 3. (a) shows a relatively narrow opening between the epiglottis and the posterior pharyngeal wall. (b) shows the vocal tract during exhale,and (c) during inhale, with the fiberscope in the same position as during the other articulations. In these photographs the tongue is so far forward as to be out of the frame



[ħ] in [niħum]

Figure 4. During the voiceless pharyngeal fricative a relatively wide opening is seen between the epiglottis and the posterior pharyngeal wall.



[ʕ] in [ʃiʕur]

Figure 5. Complete closure during the phonemically voiced fricative produces a stop on the phonetic level. The articulation resembles the gesture of swallowing to the extent that the tip of the epiglottis disappears downward and reappears abruptly.

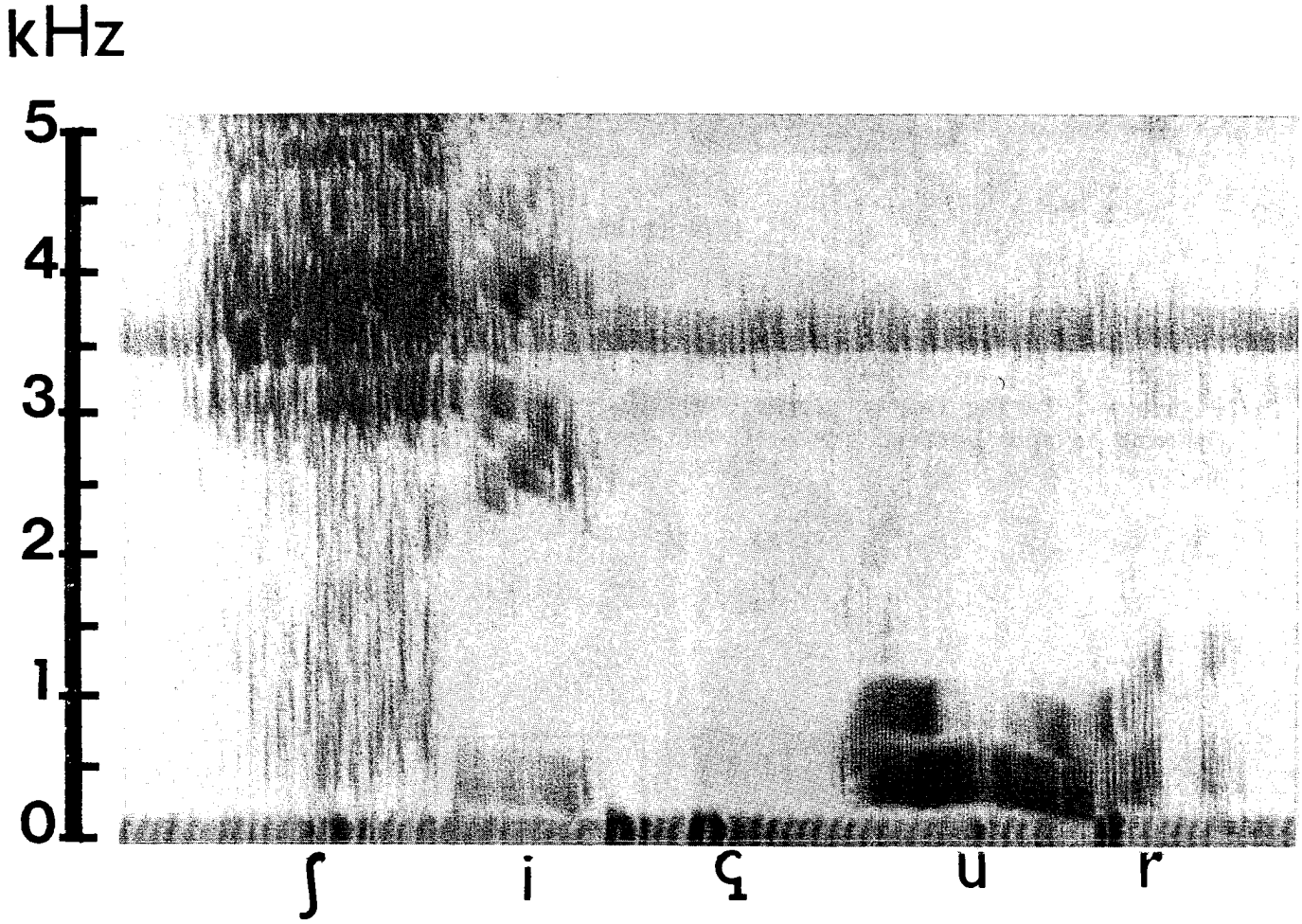
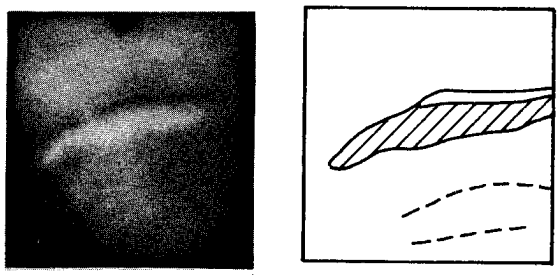


Figure 6. Phonemically voiced /ʕ/ is phonetically voiceless, as revealed by the absence of a voice bar. Here it is a stop. (The horizontal line at approximately 3.5 KHz is caused by noise from the fiberoptic light source).

glottis is one of the articulators in the production of /ħ/. We cannot see the back of the tongue in any of our data on /ħ/. The position of the tip of the fiberscope is high enough that the tongue should be clearly visible if it were significantly retracted. When the tongue does enter the field of view, for example as in /a/, it is readily identified in the picture transmitted by the fiberscope. We conclude that the tongue is not involved in the stricture of /ħ/. The stricture is made, rather, between the epiglottis and the pharyngeal wall, or the epiglottis and the apex of the arytenoids, or between the epiglottis and both the pharyngeal wall and the arytenoids. In the latter case the passage is S-shaped, and it is thus additionally turbulence-inducing.

The phonemically voiced counterpart of /ħ/ is /ʕ/. In this sound there is also a constriction between the epiglottis and the posterior pharyngeal wall. The constriction of /ʕ/ is more variable in size than the constriction observed for /ħ/; for /ʕ/ it covers a range from narrow opening to complete closure. In careful and slow pronunciation of /ʕ/ the entire width of the epiglottis touches the entire width of the pharynx, forming a complete closure. See Figure 5. From spectrograms it is seen that these sounds are stops and are furthermore phonetically completely voiceless. (See Figure 6). In more rapid and casual speech, on the other hand, the epiglottis moves back toward the posterior wall of the pharynx so there is considerable narrowing, but it can be quite wide (see Figure 7). In these cases, as seen on spectrograms, these sounds are fully voiced, and look rather like glides (see Figure 8). Between these two extremes, complete closure and a glide-like voiced continuant, there is a range of intermediate possibilities. In these we cannot see a closure between the epiglottis and the pharyngeal wall (cf. figure 9) but spectrograms show the sound to be either a voiceless stop, or sometimes to consist of creaky (glottalized) voice. In the spectrogram in figure 10 we can see creak during the pharyngeal [ʕ], but in this particular token the opening seen in the videotape is relatively large, even a little more open than the [ħ] in figure 4. Overall, we find a gradation consisting articulatorily of closure, narrow opening, wider opening, and an acoustic gradation from voiceless stop, creak, voiced fricative and glide.



[ʕ] in [ʃiʕur]

Figure 7. A narrow opening between the epiglottis and the pharyngeal wall gives as its acoustic effect a glide, in the case of this phonemically voiced fricative. For comparison see Figure 12(a), where the opening is noticeably wider for the same consonant.

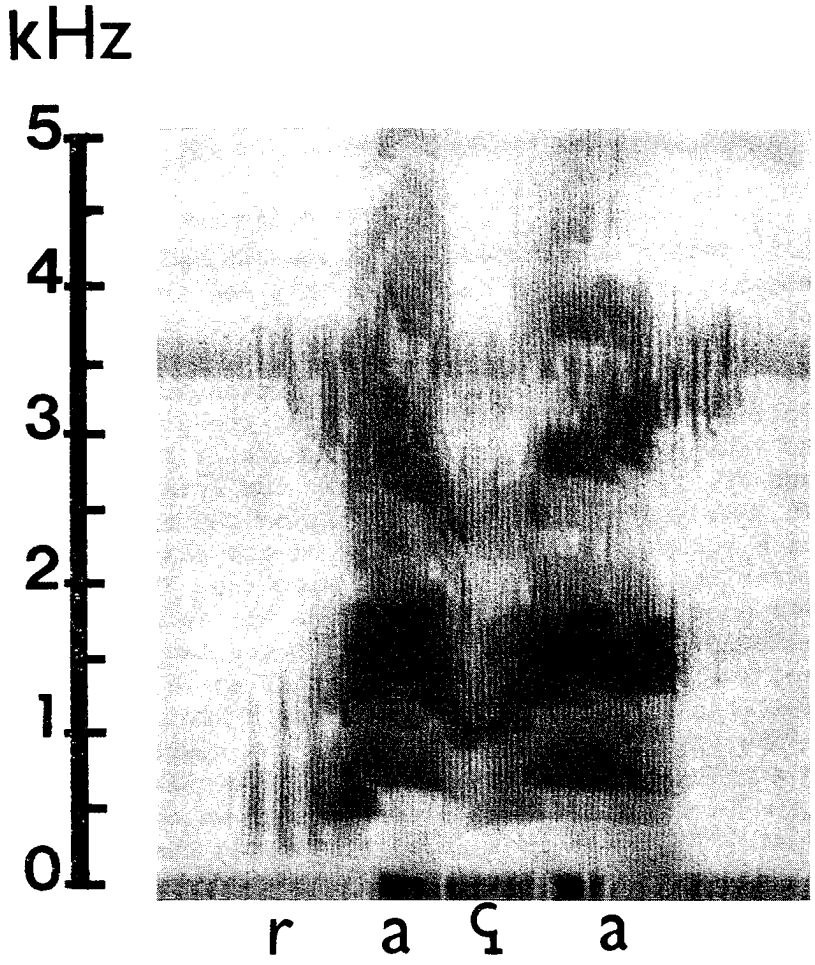
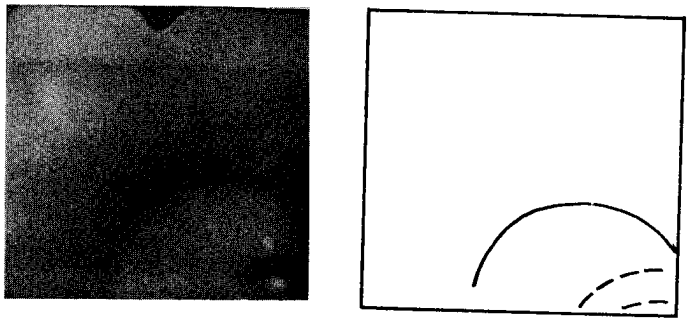


Figure 8. Rapid speech variant of [ʕ], realized as a glide resembling the adjacent [a] to each slide, particularly in the structure of the first two formants.



[ʕ] in [raʕa]

Figure 9. Narrow opening with creaky voice, as shown here, has a position of the epiglottis more nearly resembling the complete closure illustrated in Figure 5 than the narrow openings shown in the remaining figures. From this it is clear that stops cannot be detected from the fiberscopic record alone.

Our explanation of the unexpected quality of voiceless stop, where the traditional phonology leads one to expect a voiced continuant is the following: we suppose that in every case the speaker brings the vocal folds together to produce voicing. In the cases where the epiglottis moves only as close to the posterior pharyngeal wall as the position illustrated in Figure 7 (which we term "narrow opening"). A voiceless stop variant of /ʕ/ may be produced when the epiglottis moves far enough back to touch the posterior wall, or when the base of the epiglottis touches the apex of the arytenoids, or it may even be that there is a double closure epiglottis-pharynx and epiglottis-arytenoids. In any case, the air space below the closure is too small to accept enough air to allow voicing to continue for the duration of the closure. We also assume that when the epiglottis goes far enough back it also goes at the same time down far enough to touch the arytenoids (a kind of folding movement, probably a controlled, non-reflexive variant of the movement observed in swallowing); it touches, or if folded and retracted further, pushes on the tops of the arytenoids. (To put it more exactly, it pushes on the corniculate cartilages of Santorini which are immediately superior to and attached to the arytenoids and form the apex seen as a bump in the mucous membrane at the end of the vocal folds, as viewed in the fiberscope.) This pushing on the arytenoids presumably disturbs the fine adjustment needed for voicing and produces an involuntary glottal stop or, if not a stop, then glottalized voice (creak).

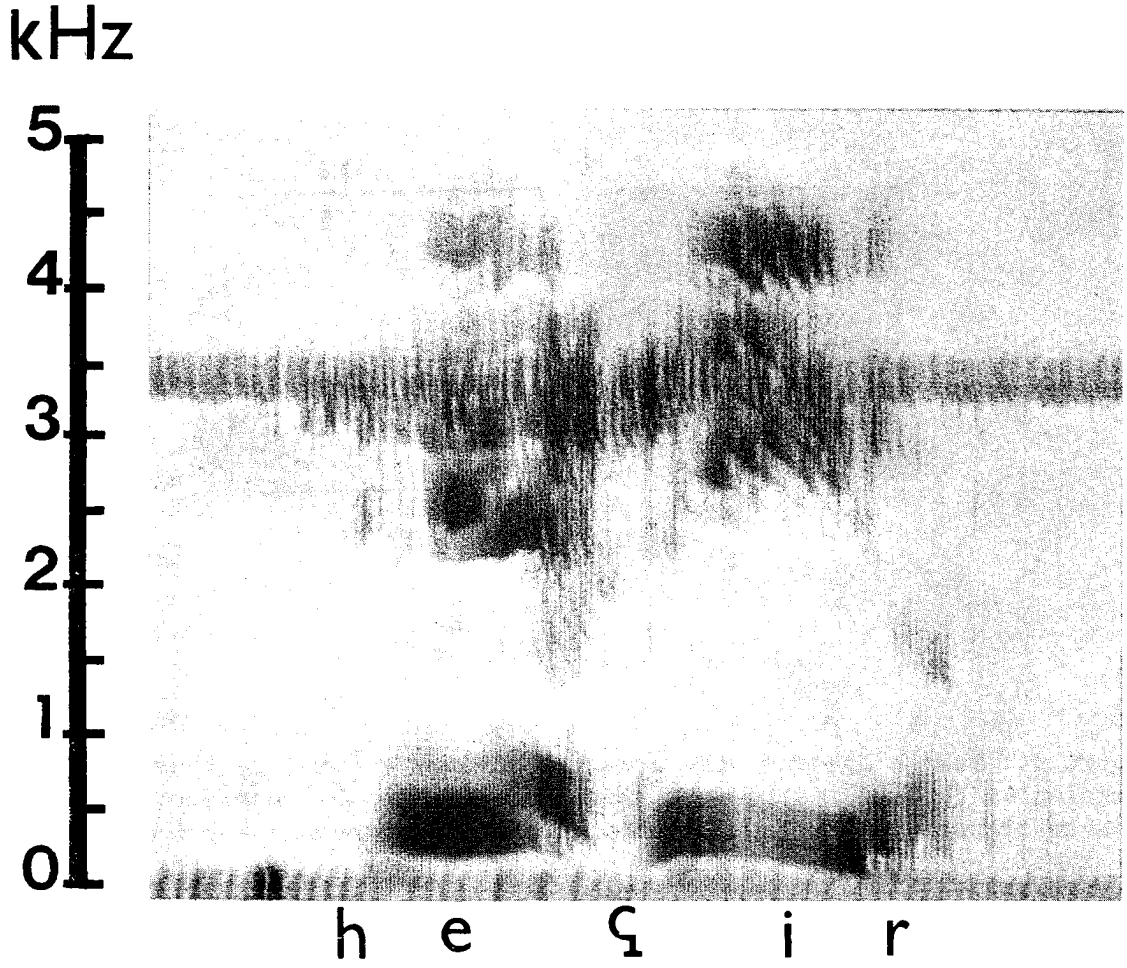


Figure 10. Creaky-voice realization of [ʕ].

To conclude we can say that phonemically /ʕ/ is voiced, and the vocal folds are always in the shape to produce voicing, but under certain circumstances, when /ʕ/ is a stop, voicing cannot be realized. /ʕ/ is produced as a glide when there is a wide constriction between the top of the epiglottis and the pharynx wall. A pharyngeal voiced fricative is made when the constriction is smaller. A pharyngeal with creaky voice is produced when the epiglottis creates an even narrower constriction with the pharynx wall, and the epiglottis base touches the arytenoids with light pressure. Pharyngeal voiceless stops are produced when the vocal folds are together and when there is complete closure in the arytenoids with or without complete closure in the pharynx. In this case we assume that the bottom of the epiglottis is touching the tops of the arytenoids with the greatest pressure.

Anatomical considerations may explain involuntary glottal stop and creaky voice connected with the production of pharyngeals. The muscles that pull the epiglottis down and back (fold it at approximately the middle of the cartilage) are the thyroepiglottic muscle and the aryepiglottic muscle. The thyroepiglotticus pulls the edge of the epiglottis down and back and possibly somewhat laterally, while the aryepiglotticus pulls it down and back and medially; this same muscle continues onto the larynx as the oblique interarytenoid. Contraction of the aryepiglotticus not only folds the epiglottis down, but the combined action of these two muscles can act to pull the epiglottis and the arytenoids toward each other. In the production of one of our subjects (Z) when the epiglottis is far enough from the pharyngeal wall to allow one to see the arytenoids, in both /ħ/ and /ʕ/, the arytenoid cartilages are seen to move forward toward the base of the epiglottis at the beginning of these consonants and back to their usual position for normal voice at the end of /ħ,ʕ/. Normally, when a muscle between two structures contracts, one structure is fixed in position and the other moves relative to it. The fact that the epiglottis and the larynx both move in this one case suggests that this same relatively unusual maneuver may occur in some percentage of the population at large. In the other eight cases the force of the muscular contraction is fully on the arytenoids, tending to pull them more tightly together than they would be if the epiglottis were not involved. This leads to creak, and also to what we referred to earlier as an involuntary glottal stop. The action of forming an epiglottic fricative or stop uses to some extent the same gestures as does swallowing. The difference between the linguistic use of these actions and the vegetative use lies in the more controlled and precise positioning in the linguistic use; the general effect of the linguistic use is less abrupt and the motions less large than in swallowing.

We find support for this concept of the epiglottis as independently controllable in a finding reported by Zemlin (1968). Although he holds the opinion that "the epiglottis contributes very little to the production of speech" (p. 126), he reports an observation that is relevant to our study. "While engaged in high-speed photography of the larynx, the author has witnessed changes in the concavity of the posterior surface of the epiglottis as the subject for photography underwent changes in the pitch of phonation. Thus, it seems as if the aryepiglottic muscle fibers actively engage in modification of the shape of the epiglottis during phonation at various pitches" (pp. 126-127).

We have reason to believe that the epiglottis is an articulator in Arabic as well as in Hebrew. We base this conclusion on our reading of the investigations of Arabic reported by Al-Ani (1970), although gaps in his presentation make it difficult to be certain of the extent to which his findings are the same as our own.

Al-Ani found that the Arabic /ʕ/ was pronounced as a voiceless stop, a voiced fricative, and a glide. "The /ʕ/ is described as a voiced pharyngeal fricative in all previous works on Arabic, literary as well as dialectal. However, after a thorough acoustical analysis, the author has found that the most common allophone of /ʕ/ is actually a *voiceless stop* and *not* a voiced fricative." (1970, 62). [Italics original.] "Oftentimes the space that /ʕ/ occupies appears as an irregular random striation of voiced noise with no clear tracing of formants--especially in the center of the position of the /ʕ/." (p. 63). This we take as evidence that his subjects were producing a fricative. He also says "/ʕ/ intervocalically [may appear] as a glide continuation of the preceding and following vowel formants" (1970, 63).

Our interpretation of the spectrograms Al-Ani presents fully supports his conclusions. Unfortunately, however, the X-ray tracings he presents do not show the epiglottis, the articulations of which might explain his spectrographic findings, even though he says that the X-rays themselves are "extremely clear and cover the whole vocal tract, lips to glottis" (p. 59). We see that he is aware of the problem in his tracings, namely that they show no occlusion, even for sounds that spectrograms show to be stops,when he says that the "tongue positions in producing the pharyngeals and glottals are quite clear but, unfortunately, this is not enough" (p. 59). He suggests in a footnote that a vertical axis X-ray would clear up the problem (p. 59, 3), but omits consideration of the epiglottis as a constrictor, and attributes the constriction of the voiceless pharyngeal /ħ/ (and, we would suppose, probably the corresponding voiced /ʕ/) instead to the tongue: "In producing the /ħ/, a constriction is formed by the dorsum of the tongue against the posterior pharynx" (p. 60), although his tracings give minimal support, if any, to this interpretation of a constriction. Unfortunately, his tracings show the tongue only, and provide no indication of epiglottis position.

The remaining point we want to discuss with respect to Al-Ani's findings compared with ours is the difference in frequency of the stop variant of /ʕ/. Al-Ani found the stop to be the commonest pronunciation, we found it to be the least common. In our data the stops occur only in slow careful speech,and we think it must be the same in Al-Ani's sample. The bulk of the material he deals with is spoken in a very slow or lento style, as we see from his spectrograms. Most of these spectrograms are of single syllables or single words, and most of these short samples occupy between 600 and 1000 ms.

Overall, we take the findings of Al-Ani as showing that in Arabic, as in Hebrew, there may be a pharyngeal stop as a realization of /ʕ/, and that stop is at least not made by the tongue and the pharynx. One reason to believe that Arabic pharyngeals (as well as pharyngealized sounds) may be articulated by the epiglottis is that some of our Hebrew subjects also speak Arabic. Proof of this suggestion will have to await an opportunity to study Arabic by fiberscopic methods.

To summarize: with the pharyngeals we have no doubt that the epiglottis is an independently moved articulator. In the extreme position of /ʕ/, that is, in the cases where the epiglottis folds down so that the tip disappears momentarily and then pops back up out of the throat, the epiglottis is clearly making an independent movement, pulled by the contraction of the aryepiglottic muscles, and not merely being pushed back by the tongue. Properly speaking, perhaps we should call /ħ,ʕ/ epiglotto-pharyngeal or even epiglotto-arytenoidals, but because as far as we are aware the epiglottis is the only articulator found in the pharynx, we will continue to refer to these sounds as pharyngeals.

(2) *The vowel /a/*

The vowel /a/ shows a clear relationship in its production and in its acoustics to the pronunciation of a glided /ʕ/. In production the opening between the epiglottis and the pharyngeal wall is similar in shape to that of pharyngeals, the principal difference being in the size of the opening, which is larger for /a/.

Another difference between the role of the epiglottis in the production of /a/ and /ʕ/ or /ħ/ seems to lie in the degree of independence of the epiglottis from the tongue. For the pharyngeals it is clear that the epiglottis folds down into the pharynx to make contact with the lateral and posterior walls of the pharynx; in these articulations the entire anterior surface of the epiglottis and the glosso-epiglottic fossa is seen, while the tongue remains forward, out of the field of view of the fiberscope. In the case of /a/, however, such folding may or may not be seen. When folding is present, the base of the tongue is not visible. In other cases, the epiglottis does not appear to fold, and may be pushed back toward the pharynx by the root of the tongue: then the articulation may be between the epiglottis and the pharynx more because the epiglottis happens to be between the tongue and the pharyngeal wall than because the

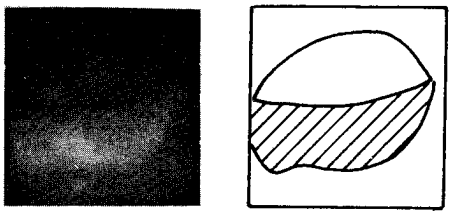
epiglottis is making an independent articulation. Nevertheless, in all cases of the production of /a/ we see the lateral edges of the epiglottis pressed firmly against the pharyngeal wall, and the lateral walls of the pharynx touching the epiglottis. The articulation is thus in all cases between the epiglottis and the pharynx (and never between the tongue and the pharynx). See Figure 11. That is, the escape of air is always between the epiglottis and the pharynx (i.e.medial to the edges of the pharynx) and never between the tongue and the pharynx (i.e. never lateral to the edges of the epiglottis).

To take a specific example of the use of the epiglottis in /a/: in the minimal pair /raʔa/, /raʕa/ ('see', 'shepherd') we see that the shape of the opening is identical during /a/ and /ʕ/, that the anterior portion of the opening is formed solely by the epiglottis and not by the tongue, that the lateral walls of the pharynx close firmly against the lateral edges of the epiglottis, and that the point of maximum constriction is thus between the epiglottis and the pharynx, not between the tongue and the pharynx.[1] The difference between /a/ and /ʕ/ is that for /a/ the opening is about three times as large in the anterior-posterior dimension and about twice as wide as /ʕ/, in one subject. The same relation holds for other subjects, although exact proportions vary somewhat . See Figure 12.

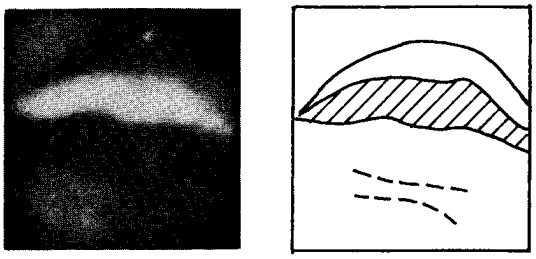
We find support for our interpretation of the production of /a/ in the findings of Russell (1931). Concerning the vowel [ɑ], Russell (1931, 39), based on extensive lateral X-rays, concluded that the shape of the tongue was not important: in order to produce this vowel there need only be a narrowing between the epiglottis and the wall of the pharynx. His lateral x-rays could only show that the tip of the epiglottis was consistently close to the posterior pharyngeal wall during /a/. Our fiberscopic study supplements his findings with the discovery that the lateral pharyngeal wall also contracts, making a relatively sphincter-like constriction. This near-closure is the point of greatest supra-glottal narrowing of the vocal tract during /a/ and is thus shown to be the acoustically significant constriction.

Additional support from the field of medicine is reported in Laufer (1977, 124, footnote 8) where he found further evidence that the epiglottis was consistently involved in the production of /a/, in discussions with the senior ear-nose-and-throat specialists at two hospitals in Israel. In order to perform laryngoscopies they ask patients (the native languages of whom were varied, only some of them being Hebrew speakers) to pronounce the vowels [i] or [e]. The physician then pulls the tongue down from its normally high position in the mouth during [i] or [e] by hand. The vowel [ɑ] of course leaves the oral cavity fully unobstructed, but it cannot be used in laryngoscopy because according to the physicians, the epiglottis consistently constricts the pharynx and blocks the view of the larynx. From

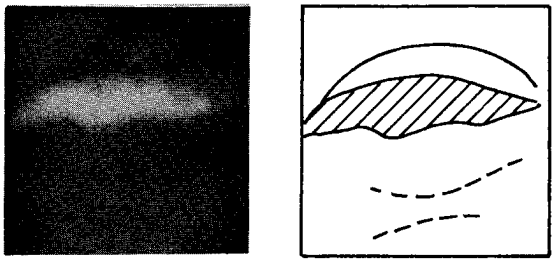
[1] Glottal stop, of course, has no effect on the articulation of /a/. We see the same effect with any /a/, as long as the adjacent consonant is not a pharyngeal or pharyngealized. During the production of /ʔ/ the epiglottis does not move.



(a) [a] in [ra?a]

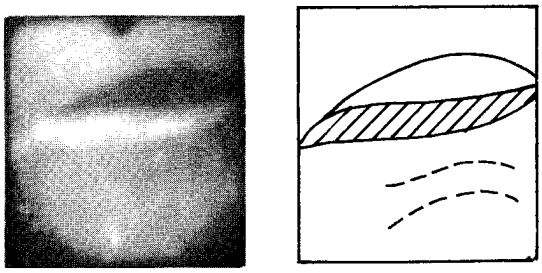


(b) [a] in [ra?a]

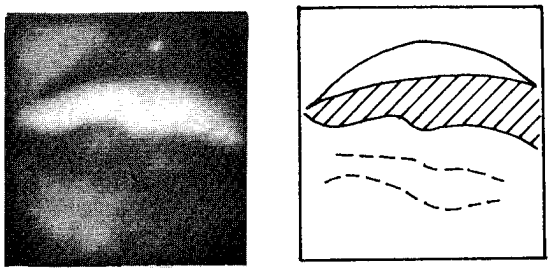


(c) [a] in [ra?a]

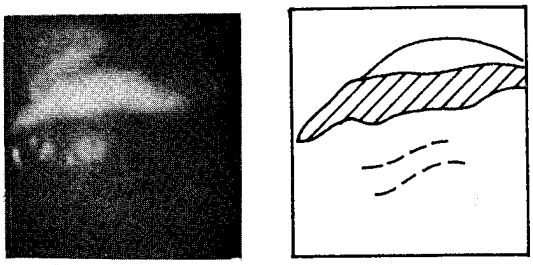
Figure 11. The epiglottis assumes different positions in different subjects for the vowel /a/. (a) is Subject S, (b) and (c) are Subject Z. However, in all cases the position resembles that of the phonemically voiced fricative /ʕ/ in the commonest variant of that phoneme.



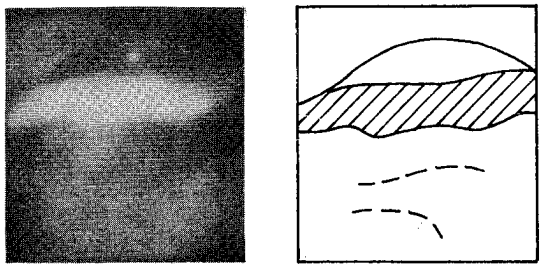
(a) [ʕ] in [raʕa]



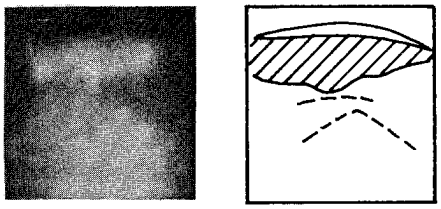
(b) [a] in [raʔa]



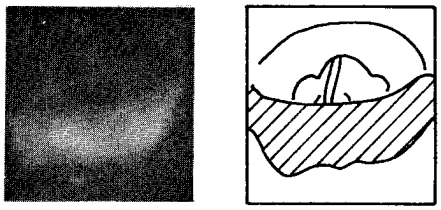
(c) [ʕ] in [raʕa]



(d) [a] in [raʔa]



(e) [ʕ] in [raʕa]



(f) [a] in [raʔa]

Figure 12. Comparison of size and shape of opening in /a/ and the phonemically voiced fricative /ʕ/. (a-d) are Subject Z, (e-f) are Subject S. (a, c, e) are the consonant, (b, d, f) are the vowel.

these facts we infer that the epiglottis is involved in the production of of [ɑ], in many (and we believe possibly all) languages.

As further support for the relation between pharyngeals and /a/ we note that in Hebrew there is a phonological rule that vowels adjacent to pharyngeals usually become /a/ (the exceptions being unimportant for the purposes of this paper). This change can be explained as an assimilation in epiglottis position: the epiglottis in proximity to the posterior pharyngeal wall forms the point of maximum constriction in the pharynx for both [a] and pharyngeal consonants.

(3) *Whisper*

Only five of the nine subjects actually produced whisper; the other four either did not read the text in whisper, or they substituted murmur for whisper. We observe that murmur and normal voice are similar, and whisper differs from them: taking the position of the epiglottis in voice and murmur as the normal position, the position during whisper may be described as more retracted. Without further evidence, it is difficult to be sure of the function of this retracted position in whisper. We hypothesize that it may contribute to the production of turbulence, beyond that which is produced at the glottis, this additional turbulence coming from friction between the arytenoids and the base of the epiglottis. We see this as related to our hypothesis regarding the role of the base of the epiglottis and the arytenoids in some cases of /ħ/ in producing supraglottal turbulence.

SUMMARY

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

Appendix I. Glossary

Transcription	Gloss	Figures
[ʔaher]	'another'	1
[heʕir]	'to wake s.o. up'	10
[koħo]	'his strength'	3
[niħum]	'console'	4
[raʔa]	'saw'	3, 11, 12
[raʕa]	'shepherd'	3, 8, 9, 12
[ʃiʕur]	'lesson'	5, 6, 7

Appendix II: Equipment

Data Recording

Cold light supply--Olympus CLS
Fiberscope--Olympus 4F, 4A
Video Camera--Sony color with JVC GA/20 adapter, JVC camera control CC48000
Video recorder--Panasonic VTR NV 3130 (1/2" open reel); JVC color cassette
Monitors--Unimedia 12" color, Sony Trinitron
Microphone--Sennheiser MD 41 U.

Data Analysis

Video--as above
Spectrograph--Kay 6061 A
Still photographs from video (illustrations in this paper)
Kodak K-2, medium yellow filter @f5.0
Pentax Spotmatic, no filter @f7.0

References

Al-Ani, S.H. (1970) Arabic Phonology: An Acoustical and Physiological Investigation. The Hague: Mouton.

Brosnahan, L.F. and Malmberg, B. (1970) Introduction to Phonetics. London: Cambridge University Press.

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

Chapman, W.H. (1973) Introduction to Practical Phonetics. Horsleys Green, England: Summer Institute of Linguistics.

Hardcastle, W.J. (1976) Physiology of Speech Production: An Introduction for Speech Scientists. London: Academic Press.

Heffner, R-M.S. (1964) General Phonetics. Madison: The University of Wisconsin Press.

Hockett, C.F. (1958) A Course in Modern Linguistics. New York: Macmillan Company.

Ladefoged, P. (1971) Preliminaries to Linguistic Phonetics. Chicago: University of Chicago Press.

Laufer, A. (1977) "Phonetic description of vowels". Leshonenu 41. (Jerusalem)41. 117-143 (in Hebrew).

Malmberg, B. (1963) Phonetics. New York: Dover.

O'Connor, J.D. (1973) Phonetics. Harmondsworth, England: Penguin.

Russell, O.G. (1931) Speech and Voice, New York, Macmillan.

Traill, A. (1978) Phonology of !xõ, Public Lecture, UCLA.

Zemlin, W.R. (1968) Speech and Hearing Science. Anatomy and Physiology. New Jersey: Prentice-Hall.

Acknowledgements

We acknowledge with thanks the support of the Phonetics Laboratory of the Department of Linguistics, UCLA, Peter Ladefoged, Director; and the assistance of Barry Goldstein with dissections, Willie Martin and James Heaton with photography; we also thank our 9 subjects for participating. This work was supported in part by the Faculty of Humanities of the Hebrew University of Jerusalem, and by the University of Hawaii Foundation.

Tone spacing : evidence from bilingual speakers.

Ian Maddieson

[Paper presented at the 95th Meeting of the Acoustical Society of America, Providence, Rhode Island; May 1978]

In his book *Tone Languages* Kenneth Pike (1948) suggested that level tones are maximally separated in the available tone space. A similar view is implied by Wang (1967) in his paper on tone features, and an assumption of maximal separation was made in the input to Hombert's model of tone systems (1978).[1] Figure 1 illustrates the comparative spacing of tones that would be found in languages with from 2 to 5 level tones if tones are maximally separated. If tones are maximally dispersed, the tones of a language with only 2 level tones will be as far apart paradigmatically as the highest and lowest tones of a language with 3 level tones, and so on.

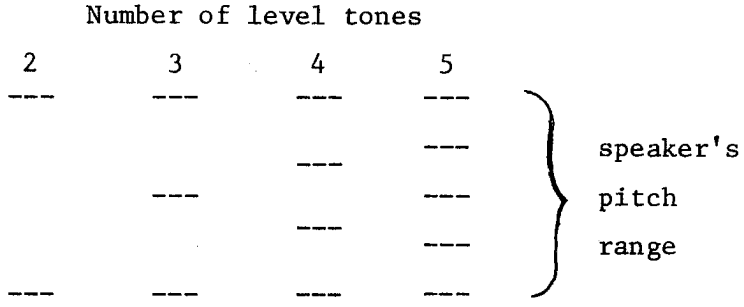


Figure 1. Maximal separation of tones.

There is, however, evidence from the way that tones are treated in loan phonology that tones are not maximally separated (Maddieson 1977). This evidence is obtained when words are borrowed from a language with fewer tones into one with more tones. For example, a word from a 2-level language with a high-low tone sequence is likely to be borrowed into a language with 3 levels with a high-mid or mid-low sequence, rather than the high-low pattern predicted by the maximal separation hypothesis. If this pattern of phonetic correspondences in tone spacing is general, the tones in languages with from 2 to 5 level tones should be compared more as shown in Figure 2, which expresses the hypothesis that the interval between any pair of most-similar level tones in a language is roughly constant.

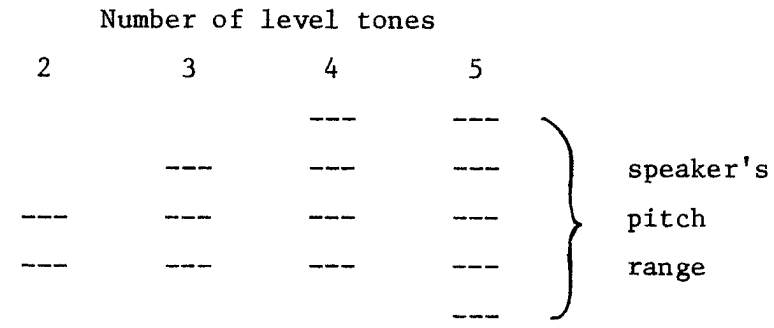


Figure 2. Constant separation of tones.

Of course, the actual pitch of a tone in an utterance is subject to influence from many factors, including characteristics of the speaker, the communicative situation, the duration of the utterance, the segmental and tonal environment, etc. In published studies , differences in measurement procedures also affect the pitch values obtained for tones. Thus, while comparing published data on intervals between tones is suggestive (Maddieson 1978), there are too many uncontrolled variables in such studies for really reliable conclusions on tone spacing to be drawn.

To permit a study of tone spacing in which as many as possible of these variables are controlled, data was obtained from bilingual speakers who spoke two tone languages. The subjects were 5 adult male African teachers or students who spoke (at least) one language with 2 level tones and one language with 3 level tones and had learned both these languages in relatively early childhood. Although one language was still considered the 'mother tongue' as the main language of the parental home, in several instances the subject was now more at ease in the other of his languages. All subjects also knew English.

Minimal or near-minimal sets of words containing the contrasting tones of the languages involved were selected. These were matched for segmental composition as far as possible across each pair of languages spoken by any subject. The words were then embedded in similar positions in sentences of equal length and similar tone pattern in the two languages. A reading list of sentences for each language was then prepared containing additional sentences and the subject was rehearsed in reading the relevant lists. Each subject then read each of the two appropriate lists a number of times under good acoustic conditions while a recording was made. Narrow-band spectrograms were made of 6 tokens of each utterance and the pitch at the mid-point of the vowel was calculated from the frequency of the highest clearly visible harmonic. The results are summarized in Table 1, which lists the words used for each subject, the pitch in Hz of the tones on these words and the differences between tones rounded to the nearest whole number.

The 5 comparisons in Table 1 show that the difference between the tones in the 2-level language is significantly smaller than the difference between the high and low tones in the 3-level language. Except in the case of the Akan/Adangme pairing, the interval in the 2-level language is on the order of half the size of the interval in the 3-level language.[2] It is reasonable to posit a more or less constant interval between a pair of most-similar tones for a given speaker and explain this relationship as resulting from the use of two such intervals in the 3-level language. Note that the size of such an interval is markedly speaker-dependent; subject IK has an unusually low voice with a narrow range but has a consistent interval between adjacent tones of about half the size of the other subjects.

Two further sets of data are presented in Table 2. In the first of these, words not segmentally matched are compared. The results are very similar to the matched data from the same subject MEI speaking Edo and Yoruba. In the second, segmentally matched words were used but they were in tonally contrastive environments in one of the languages. The comparison is between Akan and Adangme (subject JN) but in the Adangme sentence the

measured word /ba/ was followed by a high tone when it had a low tone and by a low tone otherwise. The effect of this tonal context on the measured low tone may have been sufficient to reduce the difference between high and low that would ordinarily have been observed. As it stands, however, this set of data would be more accurately described by a maximal separation hypothesis. But note that it is atypical and differs from the better matched set of data from the same subject in Table 1.

In general, our findings suggest that the hypothesis of constant separation of tones is correct, at least for these subjects and these languages. It remains to seen if it will continue to hold for comparisons involving systems with 4 or 5 level tones, or for tone languages from other parts of the world.

Acknowledgments: This work would not have been possible without the patient and generous assistance of the five subjects involved to whom I owe a large debt of gratitude. I was also greatly assisted by Linda Hunter Dresel of the University of Wisconsin, Madison to whom special thanks. The work was supported by a grant from the NSF to the UCLA Phonetics Laboratory.

Footnotes.

[1] Hombert modified this assumption in the light of the output obtained from his model. He proposes that 2 levels are not maximally separated but that 3 or more are. This 'hybrid' hypothesis deserves more study.

[2] The Akan/Adangme results seem to stand out from the remainder largely because the interval between the 2 levels in Akan is so large. It could be that Akan has effectively travelled most of the way to becoming a 3-level language through the incorporation of lexical 'downsteps' into a sufficiently large number of underlying forms. If this is so, the relevance of the Akan/Adangme pairing to this study is reduced.

References.

Hombert, J-M. (1978) "A model of tone systems" *Elements of Tone, Stress and Intonation* (D.J. Napoli, ed) Georgetown University Press, Washington D.C. : 129-143.

Maddieson, I (1977) "Tone loans: a question concerning tone spacing and a method of answering it" *UCLA Working Papers in Phonetics 36*:49-83.

Maddieson, I. (1978) "Universals of tone" *Universals of Human Language 2: Phonology* (J.H. Greenberg, ed.) Stanford University Press: 335-365.

Pike, K.L (1948) *Tone Languages* University of Michigan Press, Ann Arbor.

Subject		Language	Words	N	SD	Mean (Hz)		Differences		
MEI	(M)	Edo	á.rá	6	4.12	139				18
			à.rà	8	5.81	121				
		Yoruba	ō.rá	6	3.86	145		22		42
			ō.rō	6	2.6	123				
			ō.rà	6	9.81	103		20		
IS		Hausa	gá.dá:	6	2.03	117				22
			gà:.dà:	5	4.92	95				
	(M)	Nupe	(1)tsā.mā́	6	1.99	135		14		36
			ē.dū́	6	3.1	121		23		
			ē.dù	6	6.62	99				
IGM		Hausa	gá.dá:	6	5.13	128				17
			(2)yà:.rà:	5	4.16	111				
	(M)	Nupe	ē.dú	6	3.77	134		15		37
			ē.dū	6	3.72	120				
			ē.dù	6	2.92	97		22		
IK		Hausa	háu∫í	6	1.76	87				8
			háu∫ì	7	4.62	79				
	(M)	Jaba	(3)dzí.ī	6	3.04	103		9		19
			dzì.dzī	6	1.53	94		9		
			(4)tsì	6	2.65	85				
JN		Akan	dí	5	5.31	136				29
			dì	6	3.77	107				
	(M)	Adangme	dé	6	3.59	134		23		39
			dē	6	3.34	110		16		
			dè	5	1.27	95				

Table 1. Measurements of tones in 2- and 3-level languages for 5 bilingual speakers. Tones are marked ´ high, ¯ mid, ` low. Subjects' mother tongue is indicated by (M). Measurements were made on the word-final vowel, except in the case marked (3). Because of idiolectal differences a substitute word was measured in the cases marked (1) and (2). The original target word was not tonally distinct. In the case marked (4) a minimal contrast could not be found, so a voiceless initial consonant item with low tone was accepted.

Subject	Language	Words	N	SD	Mean (Hz)		Differences		
MEI	Edo	èní	6	5.68	143				19
		èní	10	6.3	124				
	Yoruba	Ɪgbá	6	3.7	141		20		41
		Ɪgbā	6	3.09	121				
		Ɪgbà	5	2.36	100		21		
JN	Akan	bá	6	3.69	143				35
		bà	6	3.40	107				
	Adangme	bá	6	5.84	146		22		35
		bā	6	4.33	124				
		bà	6	3.12	111		13		

Table 2. Additional measurements of tones in unmatched environments for subjects MEI and JN. Unmatched vowel qualities have very little effect on the Edo/Yoruba differences for subject MEI - Compare Table 1. However, contrasting following tones may have reduced the differences between low tone (with following high) and high and mid tones (with following low) in Adangme for subject JN; but see the comment in footnote 2.

More on the Representation of Contour Tones

Ian Maddieson

[To appear in Linguistics of the *Tibeto-Burman Area*]

In LTBA IV.I Gandour and Fromkin (1978, cf Gandour 1975) draw attention to an alternation of tone in the Tai dialect of Lue (Li 1964) which may provide crucial evidence in favor of the representation of contour tones as indivisible units rather than as sequences of level tones. They propose that the analysis of this alternation is both simpler and more insightful if unit-contour features are posited. The advantages they claim for their solution are open to question.

Briefly stated, Li reports that there are six contrastive tone patterns on Lue syllables. Five of these do not alternate. These are three level tones, transcribed [55], [33], [22], a mid-low falling contour [31], and a mid-high rising contour [25] which is represented as [35] by Gandour and Fromkin. The sixth pattern is reported as a low-mid rising contour [13], alternating with a low level [11]. The low level variant appears before the mid-high or mid-low contours, but the low-mid rising contour appears before a level tone or pause. Gandour and Fromkin suggest that the correct way of viewing this distribution is as the product of a dissimilation process: Before a contour tone a level variant appears, but before a level tone a contour variant appears. To formulate the rule describing this dissimilation process, a feature must be available to categorize tones as level or contour, e.g. as [- CONTOUR] or [+ CONTOUR]. Using the features proposed by Wang (1967), the Lue tones would be specified as in (1).

(1)

	55	33	22	31	35	13	11
HIGH	+	-	-	-	+	-	-
CENTRAL	-	+	+	-	-	-	-
MID	-	+	-	-	-	-	-
RISING	-	-	-	-	+	+	-
CONTOUR	-	-	-	+	+	+	-

If the rising variant [13] is taken as the underlying tonal value in the alternating syllables, then the rule can be written as (2).

(2) $\begin{bmatrix} - \text{ HIGH} \\ + \text{ RISING} \end{bmatrix} \longrightarrow$ [- CONTOUR] / ___ [+ CONTOUR]

As Gandour and Fromkin point out, this rule states directly and simply that a low-mid rising contour becomes level when immediately preceding another contour tone. But whether this is an insightful, or in fact an observationally adequate account of the alternation depends on some information that is lacking in Li's article. Notice that the rule (2) will change the sequence [13] [13] to [11] [13]. Li only reports that the level variant appears before [31] and [35]. Unless the low-mid rising tone itself also triggers the change to level there is no dissimilation process to be captured, and the rule would instead be as in (3).

(3) $\begin{bmatrix} - \text{ HIGH} \\ + \text{ RISING} \end{bmatrix} \longrightarrow$ [- CONTOUR] / ___ $\left\{ \begin{bmatrix} + \text{ HIGH} \\ + \text{ RISING} \end{bmatrix} \right.$ $\left. \begin{bmatrix} - \text{ RISING} \\ + \text{ CONTOUR} \end{bmatrix} \right\}$

Here the representation of contours with unit-contour features does not seem to be contributing to an insight since the class [+ CONTOUR] does not function in rule (3). Instead the environment is simply the disjunction of the mid-high rising and mid-low falling patterns.

Assuming that the distribution of the [11] and [13] variants is as (3) states, an attempt to rescue a dissimilatory explanation can be made by reversing the assumption about which of the variants is underlying. In this case Lue would have four underlying level tones and only two contours. The contour variant [13] would be derived by rule whenever a level tone followed an underlying [11] tone. Note that the rule would also have to effect the same change when a pause followed, and also that [13] [13] would be derived from the underlying sequence [11] [11]. Using the features in (1), the rule would have the form given in (4).

(4) $\begin{bmatrix} - \text{ HIGH} \\ - \text{ CENTRAL} \\ - \text{ RISING} \end{bmatrix} \longrightarrow [\ + \text{ RISING}] \ / \ \_\_\_ \ \begin{Bmatrix} [\ - \text{ CONTOUR}] \\ \text{Pause} \end{Bmatrix}$

Although the class [- CONTOUR] functions in the environment for rule (4), the rule, unlike rule (2),does not formally effect a dissimilation of the feature [CONTOUR]. However, if the feature inventory in (1) is revised by discarding the feature [RISING] in favor of the feature [FALLING] and the appropriate changes in other feature values are made, a rule which does formally represent a dissimilation can be written, namely (5).

(5) $\begin{bmatrix} - \text{ HIGH} \\ - \text{ CENTRAL} \\ - \text{ FALLING} \end{bmatrix} \longrightarrow [\ + \text{ CONTOUR}] \ / \ \_\_\_ \ \begin{Bmatrix} [\ - \text{ CONTOUR}] \\ \text{Pause} \end{Bmatrix}$

This works because, with the substituted feature, a tone which is specified [+ CONTOUR, - FALLING] is a rising tone. In (4) the feature [- RISING] must be changed to [+ RISING] because a tone which is [+ CONTOUR, - RISING] is a falling tone. There cannot be three *distinctive* features [RISING], [FALLING] and [CONTOUR] because given any two of these the third is entirely redundant, hence never distinctive. In other words, the description of the process involved in Lue tone alternation becomes an artifact of an essentially arbitrary choice between features. This is doubtful evidence to use for justification of the features involved.

In addition to the problem of arbitrariness, both of the solutions (4) and (5) require the unmotivated change of a low level tone [11] to low-mid rising [13] before pause, and consequently the unnatural disjunction of level tones and pause in the environment. These problems are not present if the distribution of the variants is as rule (2) states, i.e. the sequence [11] [13] occurs and not [13] [13]. Thus the claim that unit-contour features permit an otherwise unrepresented insight into this linguistic system to be captured depends on which of these sequences does occur. Yet this is undocumented; Li does not say which of the variants appears before [13].

Let us now turn away from assessing the difficulties in the way of accepting the unit-contour solution and consider the objections that are raised against a non-unitary analysis of the contours. Gandour and Fromkin compare their preferred solution with one using only levels. They convert Li's numerical notation directly into sequences of levels. Thus [31] is translated as a mid tone followed by a low tone, and [33] is also converted

into a sequence, of a mid tone followed by a mid tone. Because of this, the change from [13] to [11] is regarded by them as a change in the second element of a sequence from mid level to low level. Their rule, using a feature system proposed by Woo (1969), is given as (6). Note that this rule is written so that [13] is not changed before another [13] (unlike rule (2)).

(6) $$[\,-\,\text{LOW}] \longrightarrow [\,+\,\text{LOW}] \;/\; [\,+\,\text{LOW}] \_\_\_ [\,-\,\text{LOW}] \begin{Bmatrix} [\,+\,\text{HIGH}] \\ [\,+\,\text{LOW}\,] \end{Bmatrix}$$

Gandour and Fromkin dismiss (6) with the remark that "the complexity and phonetic implausibility of this rule obscure the simple tonal process underlying this alternation". In fact the rule represents a perseverative assimilation, which is not an implausible process but a very commonly encountered one. The implausibility may be thought to lie in the disjunction of [+ HIGH], [+ LOW] in the environment, but this is a consequence of the decision to use Woo's system of tone features which does not provide any way of identifying the set of 'non-mid' tones. Most alternative feature sets do provide a simple characterisation of the required set of tones, as for example [- Mid] in the features proposed by Fromkin (1972).

In preference to (6) the following account can be proposed, which seems both explanatory and formally straightforward. First, it is redundant to treat the level tones as sequences; they should be represented as single tones. The notation [55], [11] etc was developed to reflect a difference between syllable types, [55] being written to represent a high level tone in a long syllable and [5] to represent a high level tone in a short syllable (usually one with a final stop). The distinction is not one of tone, and besides is irrelevant to Lue. Now, the simplest way of viewing the alternation of tone in Lue is to say that in the level variant the second element of the low-mid sequence is deleted, i.e. [13] → [1]. Using the three features [HIGH], [LOW] and [EXTREME] (Maddieson 1970), the tones and tone sequences of Lue would be distinguished from each other as in the matrix (7). (The feature [EXTREME] is the inverse of but is preferred to a feature [MID], cf Anderson 1978).

(7)

	5	3	2	3 5	3 1	1 3	1
HIGH	+	-	-	- +	- -	- -	-
LOW	-	-	+	- -	- +	+ -	+
EXTREME	+	-	-	- +	- +	+ -	+

The alternation simplifies the four-tone sequence [1331] to [131] and [1335] to [135]. In other words two adjacent tones in the mid range coalesce when flanked by [+ EXTREME] (nonmid) tones. This is an extremely natural process which merely modifies the relative timing of a change in pitch and the articulation of the segments in the relevant syllables. This kind of process is very commonly found in tone languages (Hyman and Schuh 1974) and this occurrence is open to a functional explanation in terms of the difficulty of distinguishing basically similar patterns such as [1331] and [131] on the same number of syllables. One possible formulation of the needed rule is given as (8).

(8) $$[\,-\,\text{EXTREME}] \longrightarrow \emptyset \;/\; [\,+\,\text{EXTREME}] \_\_\_ [\,-\,\text{EXTREME}]\; [\,+\,\text{EXTREME}]$$

If [1] and not [13] occurs before another [13] then the rule is merely one which says that [13] simplifies to [1] before any tone sequence other than

itself. This can be formulated as (9).

(9) [- EXTREME] ⟶ ∅ / [+ EXTREME] ___ [α EXTREME] [-α EXTREME]

It thus seems that a straightforward and natural account of the Lue data can be given without any need to posit the unity of contours. As a theory which posits only level tones is more parsimonious than one which includes unit contours, stronger evidence than that provided by Lue should be required before accepting the unity of contours. The interesting argument that Lue displays a dissimilation with respect to the feature [CONTOUR] remains less than convincing.

References

Anderson, S.R. (1978) 'Tone features' *Tone: A Linguistic Survey* ed. V.A. Fromkin, Academic Press, New York: 133-175.

Fromkin, V.A. (1972) 'Tone features and tone rules' *Studies in African Linguistics* 3: 47-76.

Gandour, J.T. (1975) 'Evidence from Lue for contour tone features' *Pasaa* (Bangkok) 5.2: 39-52.

Gandour, J.T. and V.A. Fromkin (1978) 'On the phonological representation of contour tones' *Linguistics of the Tibeto-Burman Area* 4: 73-74.

Hyman, L.M. and R.G. Schuh (1974) 'Universals of tone rules: Evidence from West Africa' *Linguistic Inquiry* 5: 81-115.

Li, F-K. (1964) 'The phonemic structure of the Tai Lü language' *Bulletin of the Institute of History and Philology, Academia Sinica* 35: 7-14.

Maddieson, I. (1970) 'The inventory of features ....' *Research Notes* (Ibadan) 3.2/3: 3-18.

Wang, W. S-Y. (1967) 'Phonological features of tone' *International Journal of American Linguistics* 33: 83-104.

Woo, N. (1969) *Prosody and Phonology* Ph. D. dissertation, M.I.T. Available from the Linguistics Club, Indiana University, Bloomington.

(The author is supported by a grant from the NSF to the Phonetics Laboratory, University of California, Los Angeles).

Tones and tone sandhi in Shanghai: phonetic evidence and phonological analysis

Eric Zee and Ian Maddieson

0. ABSTRACT.

It has been claimed that tone sandhi in Shanghai Chinese consists only of the rightward spreading of the tone on the first syllable of a bisyllabic or a polysyllabic compound over the whole compound. To determine if the pitch contours on monosyllables and on compounds are actually similar the pitch patterns of monosyllabic words and of compounds containing two, three or four syllables have been measured for one speaker. The findings show that tone spreading will partially explain the patterns of tone contours, on the compounds, but that there are also some more arbitrary processes involved in tone sandhi in Shanghai. A phonological analysis is presented which regards contours as sequences of level tones, and tones are treated as having an 'autosegmental' association with syllabic units.

1. INTRODUCTION.

Tone sandhi in Shanghai occurs in various types of compounds which are derived by combining monosyllabic words[1]. This kind of compound formation is a highly productive process. For instance, different types of compounds may be derived by combining two or more of the following four monosyllabic words (1) in various different orders (tones are transcribed here according to the system which will be justified in later sections of this paper:

(1)	/tsɔ/	/MH/	'to illuminate'
	/ɕiɑ̃/	/MH/	'symbol; object'
	/tɕi/	/HL/	'machine'
	/ɕiɔ/	/MH/	'small'

1 Certain words which are like these compounds in form may not be amenable to a division into constituent morphemes, e.g., certain loan words, and fossilized compounds whose historical origin is forgotten. These may however be treated analogously to those compounds formed by the productive process illustrated here.

Some of the possible compounds which may be found from this small lexicon are given in (2-4).

(2) Bisyllabic compounds:

/MH/		/MH/		M M↑
/tsɔ/	+	/çiɑ̃/	----→	[tsɔ çiɑ̃]
'to illuminate'		'symbol'		'photograph'

/MH/		/MH/		M M↑
/çiɔ/	+	/tsɔ/	----→	[çiɔ tsɔ]
'small'		'to illuminate'		'portrait (photograph)'

(3) Trisyllabic compounds:

/MH/		/MH/		/HL/		M H L
/tsɔ/	+	/çiɑ̃/	+	/tçi/	----→	[tsɔ çiɑ̃ tçi]
'to illuminate'		'symbol'		'machine'		'camera'

/MH/		/MH/		/MH/		M H L
/çiɔ/	+	/tsɔ/	+	/çiɑ̃/	----→	[çiɔ tsɔ çiɑ̃]
'small'		'to illuminate'		'symbol'		'small photograph'

('↑' = tone raised)

(4) Quadrisyllabic compound:

/MH/		/MH/		/MH/		/HL/		M H M L
/çiɔ/	+	/tsɔ/	+	/çiɑ̃/	+	/tçi/	----→	[çiɔ tsɔ çiɑ̃ tçi]
'small'		'to illuminate'		'symbol'		'machine'		'small camera'

The most extensive previous discussion of tone sandhi in Shanghai seems to be that of Sherard (1972). Sherard provides some impressionistic phonetic data on the pitch contours observed on some compounds. He proposed the following two generalizations about the patterns he found. (1) The tone contour over an entire bisyllabic or polysyllabic compound is dependent on the tone type of the first syllable, although the resultant contours do not necessarily have the same shape of that of the tone on the first syllable, and (2) bisyllabic and polysyllabic compounds with a first syllable of the same tone type have the same tone contour.

Ballard (1976), basing himself on Sherard, stated that "in Shanghai apparently the only sandhi that occurs is of the type I shall call 'right spreading'. Within a phonological word (or, a compound), all syllables after the first one ignore their citation tone and take their tone from the tone of the first (leftmost) syllable by its spreading out over the whole phonological word." (p. 12) In other words, Ballard suggests that the pitch contours on compounds do necessarily have the same shape as those on their first syllables in isolation. In fact, most of the data presented by Sherard is open to such an interpretation but there is relatively little discussion of compounds of more than two syllables.

What is of especial interest in these accounts is the suggestion that the contours are treated as unitary wholes. If the monosyllabic contour is extended as a unit over the longer compound this would represent an important piece of evidence in favor of the proposition that tone contours are in fact units in some languages (see Anderson, 1978, for some discussion of the issue). Therefore the reality of such a rule in Shanghai would have an important consequence for phonological theory. Kennedy (1953) proposed a similar rule for his native Wu dialect of Tangsi but the question of the unity of contours was not explicitly raised and it cannot be resolved from the data he cites. More extensive documentation from relevant languages is needed: this paper is offered as a contribution to that process with respect to Shanghai.

The study reported here is concerned first with comparing the shapes of pitch contours on various classes of monosyllabic words with those on an extensive variety of compound words beginning with monosyllables of different classes. Exact pitch contours have been obtained so that this comparison may be based on objective criteria. Because only one speaker was used for these comparisons, appropriate caution must be used in drawing generalisations of broader scope. However, the acoustic analysis offers a solid foundation for constructing a detailed phonological account of the processes involved in deriving the tone patterns on compounds in the speech of this subject. In particular it makes it possible to determine how far a 'tone-spreading' explanation of the sandhi processes in compounds can account for the facts of the case, and helps to clarify the issue of whther the contours on monosyllables are being spread out as indivisible units.

2. TONES ON THE MONOSYLLABIC WORDS.

Five of the eight etymologically distinct 'tones' of Middle Chinese (c.600 A.D.) remain distinct in contemporary Shanghai. Although these are traditionally referred to as tones, the differences between them concern several properties of the syllable, including the pitch, the presence or absence of a final glottal stop, che presence or absence of voicing in initial

obstruents and the duration of the syllable. It is therefore more accurate to talk of five syllable types. Three of these are long syllables with contrasting pitch patterns; these syllables are open or nasal final. The remaining two syllable types are short and have a final glottal stop but contrast in pitch. These syllable types will be identified by the letters, A, B, C, D and E. The monosyllabic words in Table I show the contrast between these types. Note that initial obstruents in syllables

	SYLLABLE TYPE	MONOSYLLABIC WORD	
Long, Open	A	[pî]	'edge'
	B	[pī]	'change'
	C	[bǐ]	'skin'
Short, Closed	D	[pɪ́?]	'pen'
	E	[bɪ̀?]	'other'

initial voicing with obstruents

[´ = HIGH; ¯ = MID; ` = LOW; ˆ = ˊ˘; ˇ = ˋ˘; ˬ = ˋ¯]

TABLE I. Syllable types and monosyllabic words.

of types C and E must be voiced, obstruents elsewhere are voiceless. The words in Table I were used for the acoustical analysis of pitch on monosyllabic words.

A word list containing 50 tokens of these words (5 x 10 repetitions), arranged in a random order, was prepared. Each word was placed in the carrier frame as shown below:

[ŋū	dɔ̀?	____	pʌ́?	nóŋ	tʰî ŋ]
I	read	____	give	you	listen
'I	read	____	for	you	to listen.'

The speaker used in this investigation was female native of Shanghai. She is in her late fifties and was once a high school teacher in Shanghai. Recently she immigrated into the United States via Hong Kong where she lived no more than five years. Her speech was judged by other native

Shanghai speakers, including the first author, and was considered to be standard metropolitan Shanghai. The speaker was instructed to read the word list at a normal rate of speech. The recording was performed in a sound treated room. The tape was analyzed using the PDP-12 computer at the UCLA Phonetics Lab. Fundamental frequency measurements were derived every 10 msec by the CEPSTRUM method . Duration measurements of the vowels were also made on the waveforms displayed on the computer screen. Because the long syllables differ substantially in duration the data from the pitch analysis for these syllables was normalized in the time domain in the following manner. The measurements of the fundamental frequency for each token were divided into five sections, each containing as nearly as possible an equal number of 10 msec intervals and the means of the values of the 10 msec intervals within each section were calculated. A mean value for each of the five sections of each token was thus obtained. These values were then averaged across tokens of the same word to obtain means for each word. These mean values are represented by the points shown in Figure 1. No such normalization process has been applied to the pitch data for the short syllables, as the duration of these syllable types is so short that normalization would severely distort the true picture of the pitch contour.

The phonetic pitch shapes of the five syllable types are shown in Figures 1 and 2. Each contour on Figure 2 represents one single token. Duration is not normalized on Figure 2. Figure 1, however, shows the pitch of syllable types A, B and C of normalized duration, and each is the average of 10 tokens. Figure 3 shows the pitch contours of three additional tokens of type D and E syllables. From these figures, we see that a type A syllable has a high falling pitch contour. The pitch contour of a type B syllable is fairly level, but has a slightly rising movement for most of its duration. A syllable of type C has a rising contour. The pitch contour of a type D syllable starts almost as high as type A and remains at this level for a short while and then falls sharply. As for type E syllables, they rise, but then fall sharply at the end.[2]

So that comparisons of length may be made the duration (in msec) of the vowels in the five syllable types is given in Figure 4. Each value on top of the dark bars is the average of 10 tokens.

These findings generally agree with previous descriptions of the pitch contrasts that occur on the monosyllabic words. There are, however, some minor differences in the phonetic descriptions of the pitch shapes. JIANGSUSHEN HE SHANGHAISHI FANGYAN GAIKUANG (JHSFG: Synopsis of Shanghai Dialect and the Dialects in Jiangsu Province, 1960) describe the pitch contours of the five syllable types according to a 5-point numerical scale, where 5 = high, as follows:

2 Sokolov (1975) presented some measurements of the initial and final fundamental frequency contours of monosyllabic words spoken by one male and one female speaker of Shanghai. His data appear not to be inconsistent with the measurements presented here.

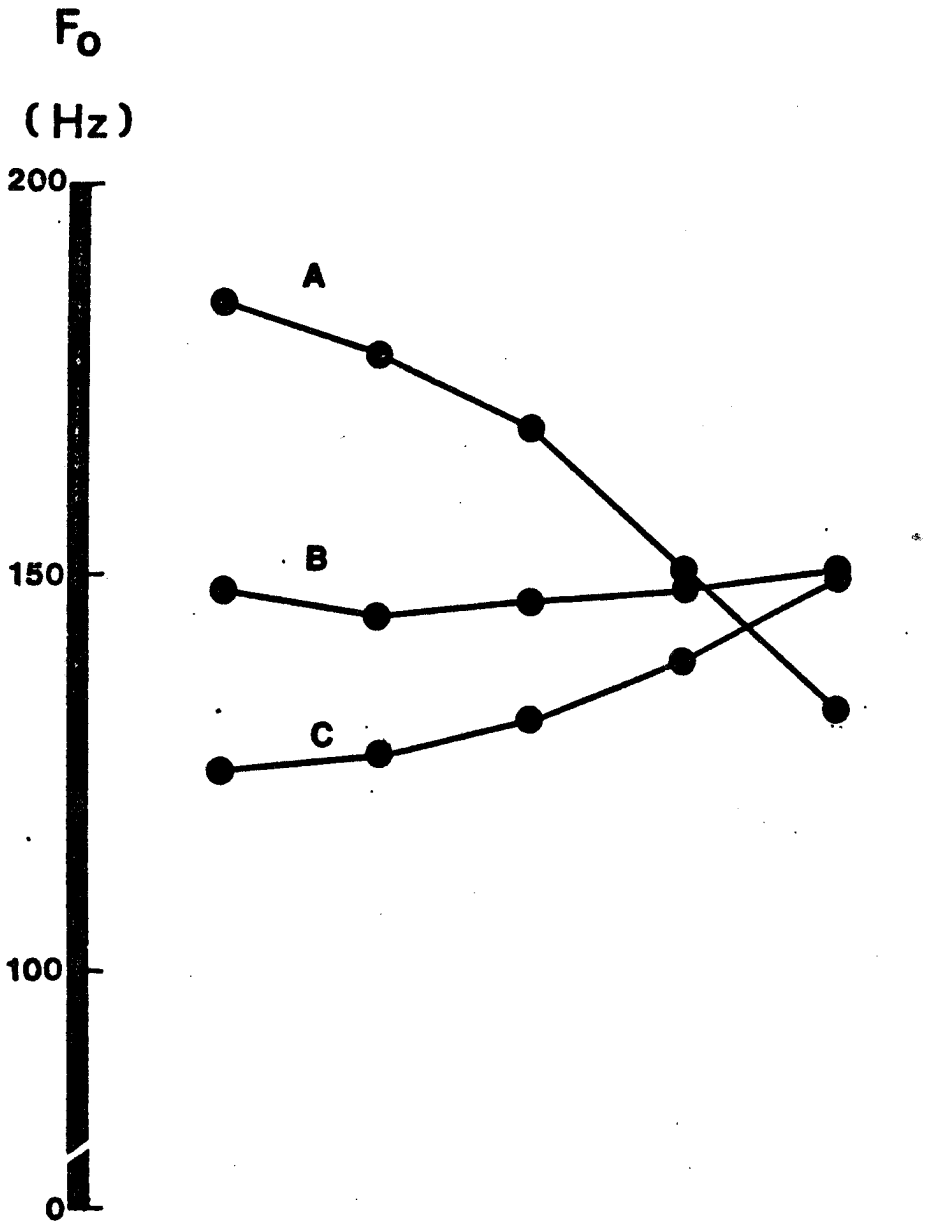


FIGURE 1. Pitch contours of normalized duration of syllable types A, B, and C.

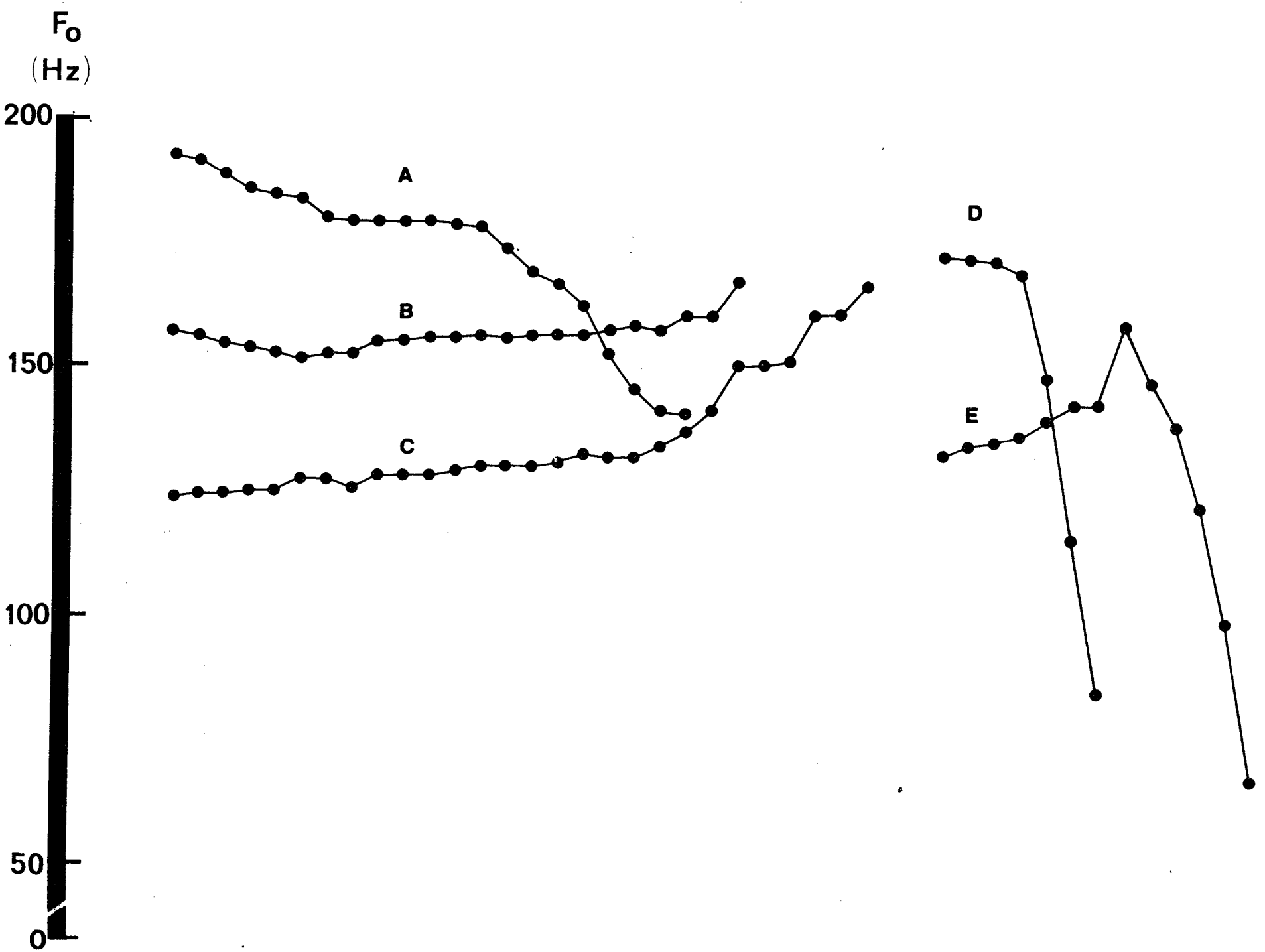


FIGURE 2. Pitch contours on syllable types A, B, C, D and E.

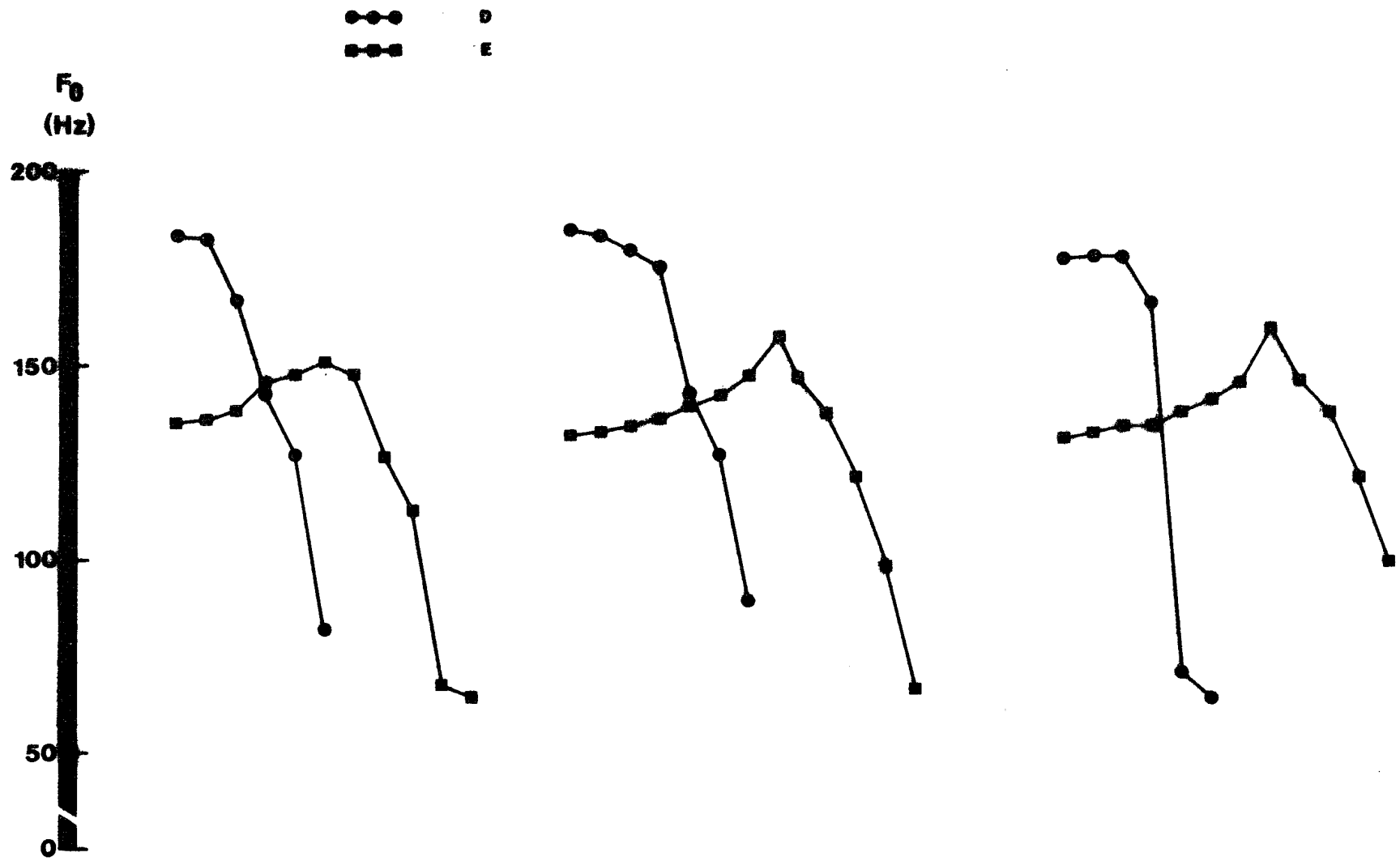


FIGURE 3. Pitch contours of syllable types D and E.

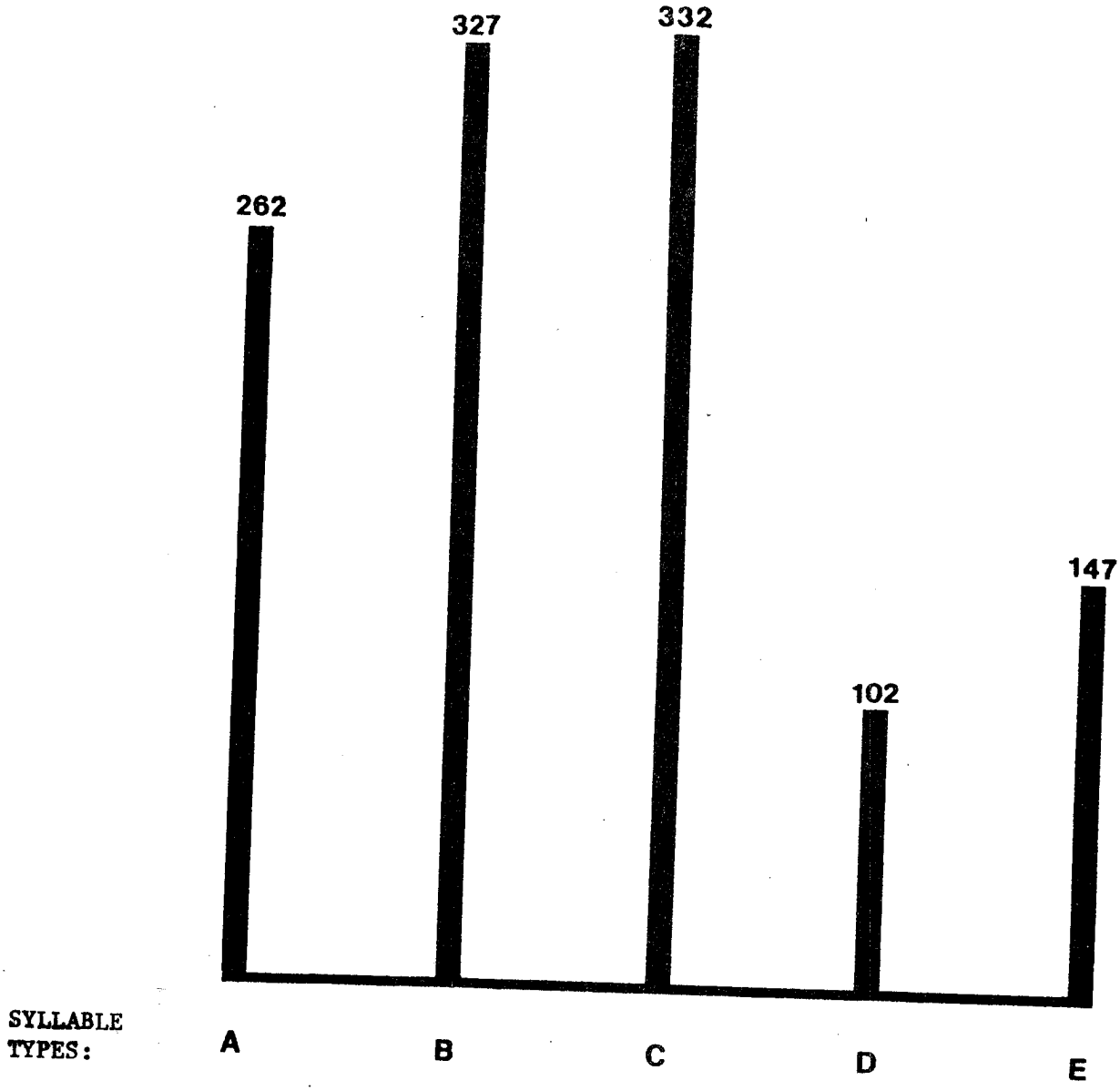


FIGURE 4. Durations (in msec) of syllable types A, B, C, D and E.

SYLLABLE TYPE		JHSFG	HFG
Long	A	53	42
	B	34	35, 434
	C	14	24
Short	D	5	5
	E	2	23

Sherard (1972) offers a qualitative description of the pitch shapes, in which syllables of type A are characterized as having a sharp falling tone, B as having a moderately high level tone with no discernible peak in pitch height in normal speech, C as having a tone in low register with clearly audible tone contour with end point higher than the onset, D as having a tone of moderately high pitch with no discernible change of pitch, and E as having a tone in low register throughout with a short rise in pitch. The differences among the impressionistic studies, as far as the pitch shape of monosyllabic words is concerned, are in the description of syllables of types B and E.

The pitch of type B syllables is described by JHSFG as mid with a slight rise (34 in numerical terms). HFG implies that it is more sharply rising (35), or is in a higher-mid range but has a dip in the middle (434). Sherard describes it as moderately high and basically level but he draws it with a slightly dipping contour. Our acoustical data show that the pitch at the onset of type B syllables is midway between the pitch at the onset of types A and C. The pitch dips slightly after the onset and then rises for most of its duration. The end point is higher in pitch than the onset. We are inclined to believe the characterization of this contour as basically mid-rising is correct. The tendency to fall after the onset which produces the dipping pattern probably can be related to the fact that basically level tones actually fall in most languages, especially on longer syllables. Apparently JHSFG did not observe this initial fall, whereas both HFG and Sherard paid some attention to it. The amount and to some extent the direction of pitch movement observed seems to vary but some rising movement is always reported.

The pitch on type E syllables has been described as either short low rising (HFG and Sherard) or as short low level tone(JHSFG). Our acoustical data (Figure 2) show that the final portion of the pitch contours of both type D and type E syllables falls to an extremely low

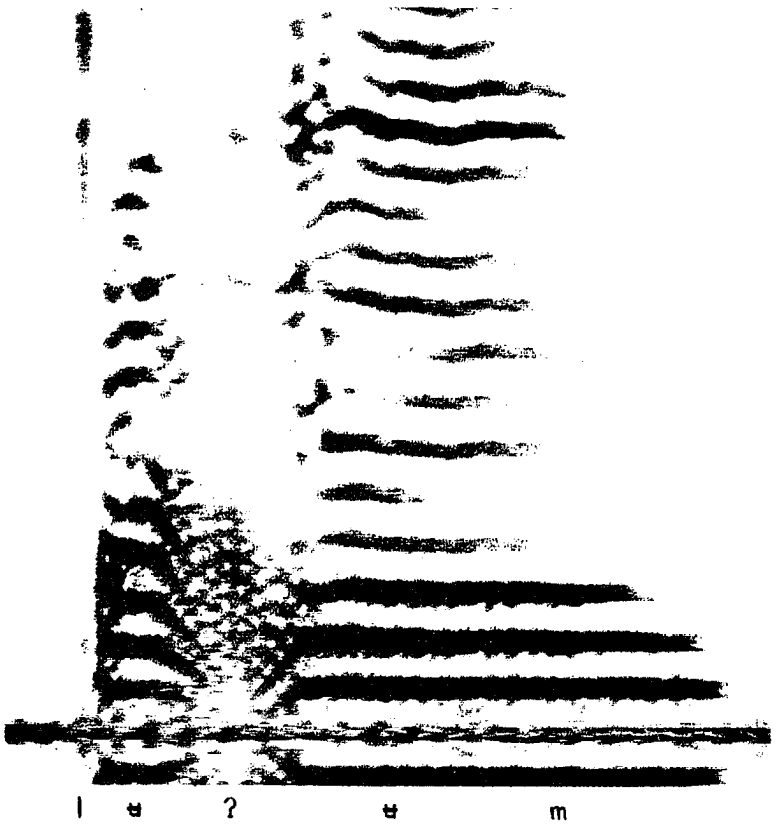
point for our subject. This sharp fall has not been reported in the previous studies, but we are doubtful if this means it was missing from the speech of their subjects. The authors may not have perceived its presence due to its extreme shortness (in our subject's speech this falling pitch movement is extremely difficult to perceive by ear). If it is noticed at all, it may rather be interpreted as a cue for the presence of the final stop. Recall that both type D and type E syllables end with a syllable final glottal stop, (C)V?. We believe that the sharp falling portion of the pitch contours is caused by the closing action of the vocal cords for the glottal stop. If this is true we would expect to find similar effects in other languages, and indeed, a similar case is found in !Xõ, a language spoken in South West Africa (Traill, in progress). In !Xõ, the final portion of the pitch contour of a vowel drops sharply when the vowel is followed by a glottal stop. Figure 5 shows a narrow band spectrogram of the pitch contour of the word [ǀʉʔʉm] "to cool down". We can see that the portion of the pitch contour that precedes the intervocalic glottal stop falls sharply.

Examining the pitch contours of type D and type E syllables in Figure 1, we see that the pitch begins to drop sharply at the 5th time point for type D syllables and at the 9th time point for type D syllables and at the 9th time point for type E syllables. We can reasonably assume that closure for the glottal stop begins to be articulated at these time points. Since we argue that the sharp fall in the pitch contours of type D and type E syllables is due to segmental influence, that is, it is a property that can be predicted in this environment, we exclude this property from the phonetic representation of the pitch contours of the short syllables. Thus,phonetically we may describe these pitch contours as high level (type D syllables) and low rising (type E syllables) for our subject. This characterisation of type D agrees with all the other sources, and this description of type E agrees with HFG and Sherard. The speech variety reported in JHSFG may have differed in having a level pitch pattern in type E syllable, but perhaps this source has merely used a simplified transcription for the two-way contrast in short syllables.

Based on our acoustical data and the above discussion,we summarize the essential pitch characteristics of monosyllables of the types A, B, C, D and E in Shanghai as follows:

(1) type A syllables - start high then fall to low.
(2) type B syllables - start at a mid level and rise slightly.
(3) type C syllables - start low then rise to the same level as the end of type B.
(4) type D syllables - high level.
(5) type E syllables - start low then rise to almost the same level as the end of type B and type C.

TYPE B/65 SONAGRAM® KAY ELEMETI



[ɪʉʔʉm] 'to cool down'

(Language: ˈXõ:)

FIGURE 5. Narrow band spectrogram of [ɪʉʔʉm].

Note that the pitch contours of type C and type E syllables are similar, but that type B starts at a higher level than C or E. This difference in the contours is much too great to be attributed to the possible phonetic effect of the initial consonants in the test items used with our subjects. The pitch levels in B and C type syllables remain sharply different even as long as 250 msec after the vowel onset, so that even though the type B word measured begins with [p] and the type C word measured begins with [b] the voicing difference is not plausibly responsible for the different contours on these syllables. (See Hombert, 1978, for more discussion on the extent of consonant perturbations of F_o).

3. PITCH CONTOURS OF THE BISYLLABIC, THE TRISYLLABIC AND THE QUADRISYLLABIC COMPOUNDS.

For comparison with the contours on monosyllables pitch contours for bisyllabic, trisyllabic and quadrisyllabic compounds were obtained using the same subject and measurement technique described above. The compounds chosen for analysis and the members within the compounds are listed in Tables II-IV. The compounds are composed of combinations of elements from the five etymologically distinct tone types in all possible combinations of two syllable types and selected sequences of three and four different types.

The possible pitch contours in bisyllabic compounds, and the pitch contours of the selected trisyllabic and the quadrisyllabic compounds are shown in Figure 6, Figure 7 and Figure 8 respectively. They are actual copies of the fundamental frequency contours displayed on the computer screen. The syllable type of each of the syllables in these compounds is indicated by A, B, C, D or E. Each dot, large or small, represents the fundamental frequency value at a certain time point, (10 msec apart). The large dots also mark the beginning and end of the pitch contour of any compound. Each contour represents one single token taken as representative of the type.

Figure 6 contains the pitch contours of the bisyllabic compounds listed in Table II. The data are arranged according to the tonal categories of the isolated monosyllables of which they are constructed. Thus, all compounds in the first row have a monosyllabic word of type A in first position, the second row have a type B syllable in first position, and so on. All the compounds in the first column have a monosyllabic word of type A in second position, the second column have a monosyllabic word of type B, and so on. As we can see, the shapes of the pitch contours of the bisyllabic compounds within each row (i. e., having the same isolation tone on the first syllable) are broadly similar, but the pitch contours are different within each column. In other words, the pitch contour of the first syllable in isolation largely determines the contour over the compound. Thus, there are five basic pitch patterns for the bisyllabic compounds. However, where the second syllable is a type D syllable and the first is

BISYLLABIC COMPOUNDS

#1 /tʰî/ 'day' + /tʰî/ 'day' ➡ [tʰítʰì] 'everyday'

#2 /tʰî/ 'sky' + /tʰɕT/ 'air' ➡ [tʰítʰɕì] 'weather'

#3 /tʰî/ 'sky' + /dǐ/ 'earth' ➡ [tʰídì] 'universe'

#4 /pî/ 'to arrange' + /tʰɕí?/ 'to compile' ➡ [pítʰɕì?] 'editor'

#5 /fî/ 'flying' + /dǐ?/ 'saucer' ➡ [fídì?] 'flying saucer'

#6 /tɕǐ/ 'to cut' + /tʂ̂/ 'knife' ➡ [tɕǐtʂ̄] 'sissors'

#7 /tɕǐ/ 'to examine' + /tʰʂ̌/ 'to discuss' ➡ [tɕǐtʰʂ̄] 'to criticize'

#8 /pǐ/ 'flat' + /dɤ̌/ 'bean' ➡ [pǐdɤ̄] 'flat bean'

#9 /tɕǐ/ 'space' + /tɕí?/ 'connection' ➡ [tɕǐtɕí?] 'indirect'

#10 /tɕǐ/ 'space between' + /dǐ?/ 'to spy' ➡ [tɕǐdǐ?] 'spy'

#11 /dǐ/ 'electricity' + /tʰî/ 'stairs' ➡ [dìtʰī] 'elevator'

#12 /dǐ/ 'electricity' + /tʰɕǐ/ 'tool' ➡ [dìtʰɕī] 'electric tool'

#13 /dǐ/ 'younger brother' + /dǐ/ 'younger brother' ➡ [dìdī] 'younger brother'

#14 /dǐ/ 'electricity' + /pí?/ 'pen' ➡ [dìpí?] 'electric heating tube for fish tank'

#15 /dǐ/ 'electricity' + /dʒǐ?/ 'pole' ➡ [dìdʒī?] 'poles (battery)'

#16 /tɕí?/ 'urgent' + /ɕŷ/ 'need' ➡ [tɕítɕý] 'urgent need'

#17 /tɕí?/ 'to connect' + /tɕǐ/ 'to aid' ➡ [tɕítɕǐ] 'to aid'

#18 /tɕí?/ 'knot' + /dɤ̌/ 'head' ➡ [tɕídɤ́] 'knot'

#19 /tɕí?/ 'urgent' + /tʰɕí?/ 'to cut' ➡ [tɕítʰɕí?] 'urgent'

#20 /tɕí?/ 'to amass' + /dʒǐ?/ 'extreme' ➡ [tɕídʒí?] 'motivated'

#21 /zʌ̌?/ 'ten' + /sɛ̂/ 'three' ➡ [zʌ̀sɛ̄] 'thirteen'

#22 /zʌ̌?/ 'day' + /tɕǐ/ 'record' ➡ [zʌ̀tɕī] 'diary'

#23 /zʌ̌?/ 'stone' + /dɤ̌/ 'head' ➡ [zʌ̀dɤ̄] 'stone'

#24 /zʌ̌?/ 'straight' + /tɕí?/ 'to connect' ➡ [zʌ̀tɕí?] 'direct'

#25 /zʌ̌?/ 'sun' + /zʌ̌?/ 'to eat slowly' ➡ [zʌ̀zʌ̄?] 'solar eclipse'

TABLE II. List of bisyllabic compounds and first and second memeber of the compounds.

(´ = H; ¯ = M; ` = L; ^ = ˆ; ˇ = ˇ; ˩ = `¯; ↑ = tone raised)

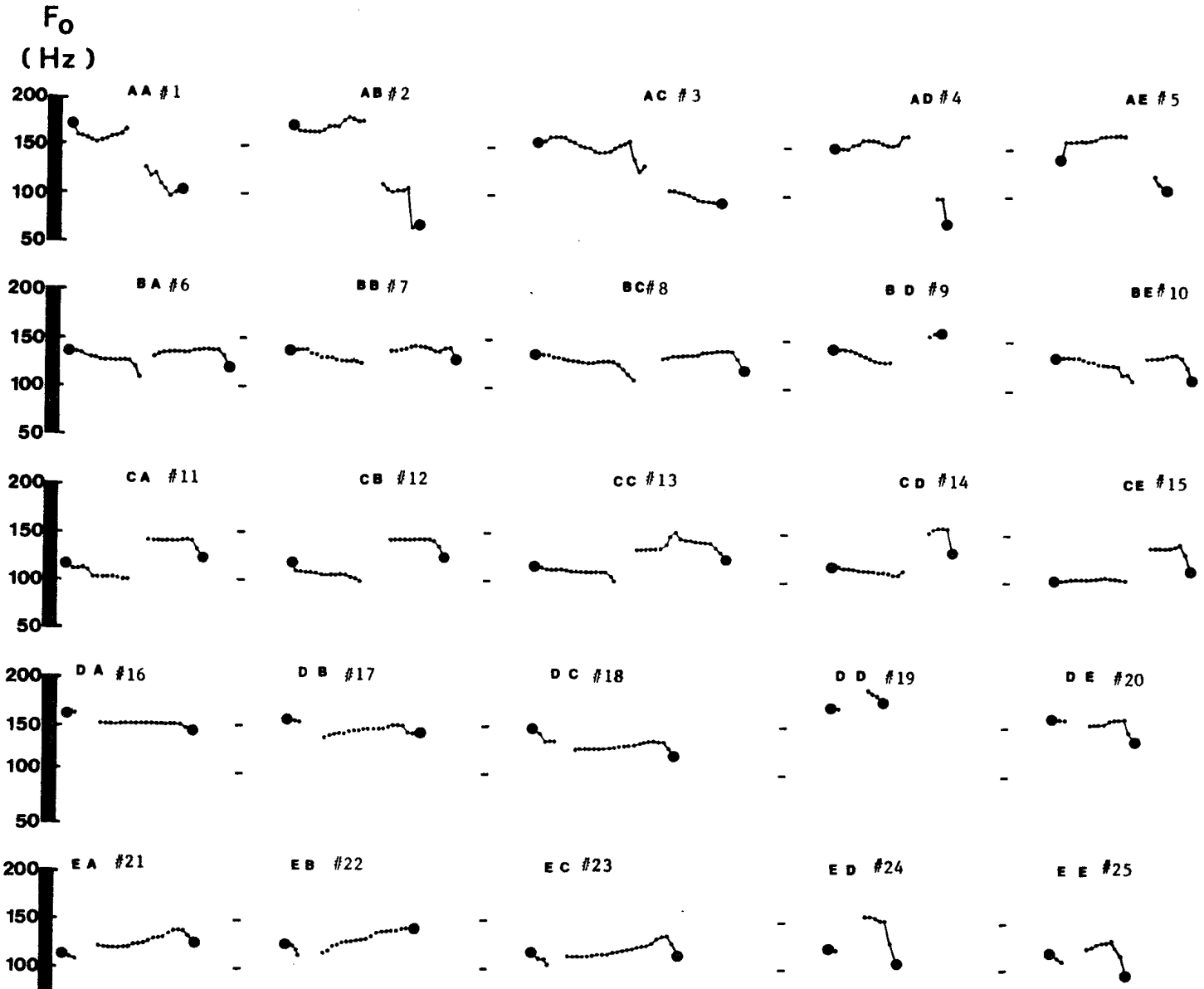


FIGURE 6. Fundamental frequency (pitch) contours of the bisyllabic compounds.

any syllable type except type A, the second portion of the pitch contour is higher than the comparable second portion of the other pitch contours in the same row. With the exception of these cases, the shapes of the pitch contours of the bisyllabic compounds are similar to the shape of the pitch contours of the monosyllabic words of the same type as their first syllable (Cf. Figures 1-3 and Figure 6). Of course, this statement is made based on the presumption that the sharp falling portion of the pitch contours of type D and type E syllables is not an intrinsic tonal property.

Figure 7 contains the pitch contours of the trisyllabic compounds listed in Table III. The data are again arranged in five rows so that all compounds in the first row have a type A syllable in first position, and in the second row have a type B syllable in the first position, and so on. The columns represent compounds with the same syllable type in second position. The third syllable may be any of the five syllable types. We can see that the pitch contours within each row are similar. Thus, there are five patterns of pitch contour that occur on the trisyllabic compounds. Notice that the shape of the pitch contour of the trisyllabic compound with type A or type E syllable as its first syllable is similar to the shape of the pitch contour of type A or type E syllable in isolation. This is however not the case for compounds with type B, C or D syllable as their first syllable. For these trisyllabic compounds, the first portion of their pitch contour is similar to the first portion of the pitch contour of their first syllable when it occurs as a monosyllabic word. The shape of the remaining portion of the contours is similar in all these cases, i.e., it falls to low.

Figure 8 contains the pitch contours of the quadrisyllabic compounds listed in Table IV. One example each is given of a quadrisyllabic compound with type A, B, C, D or E syllable as its first syllable. We can see that the shape of the pitch contour of the compound with type A syllable as its first syllable is similar to the pitch contour of a type A monosyllabic word. However, this is not so for the others. For these cases, the first portion of their contours is similar to the first portion of the pitch contour of their first syllable in isolation. The shape of the remaining portion of the pitch contours is similar in all cases. Note that the pitch pattern of first portion of the quadrisyllabic compounds with type C syllable as their first syllable is similar to the first portion of the compounds with a type E syllable as their first syllable. There are in fact only four pitch contours which occur on the quadrisyllabic compounds.

As may be seen from comparing Figure 7 and Figure 8 the shapes of the pitch contours of the trisyllabic and the quadrisyllabic compounds are generally similar. For example, the pitch contours of the trisyllabic compounds which begin with a type A syllable (the first row in Figure 7) are similar to the contour of the quadrisyllabic compound with an initial type A syllable. The exception to this observation is the compounds whose first syllable is type E. In this case the quadrisyllabic compound has a contour unlike that found on the trisyllabic compounds.

TRISYLLABIC COMPOUNDS

#1 /fî/ 'flying' + /tɕî/ 'machine' + /sɻ̂/ 'instructor' ➡ [fítɕ̄ìsɻ̀] 'pilot'

#2 /tɕî↓/ 'to exchange' + /ɕ̌ɪ̄ã/ 'sound' + /tʰɕɪó?/ 'song' ➡ [tɕíɔɕɪ̄ã̄tʰɕɪò?] 'symphony'

#3 /tʰî/ 'sky' + /v̌əŋ/ 'studies' + /dɛ̌/ 'terrace' ➡ [tʰívəŋdɛ̀] 'observatory'

#4 /sɛ̂/ 'three' + /kó?/ 'corner' + /pɛ̄/ 'board' ➡ [sɛ́kɔ̄pɛ̀] 'triangles'

#5 /jîŋ/ 'sound' + /jɪa̋?/ 'music' + /tɕî↓/ 'specialist' ➡ [jíŋjɪ̄ātɕɪà] 'musician'

#6 /tʰɕ̌ī/ 'gas' + /tʰsɔ̂/ 'cart' + /fû/ 'person' ➡ [tʰɕ̄ītʰsɔ́fù] 'chauffeur'

#7 /tsɔ̄/ 'to illumine' + /ɕ̌ɪ̄ã/ 'symbol' + /tɕî/ 'machine' ➡ [tsɔ̄ɕɪ̄ã̄tɕì] 'camera'

#8 /ǰɪ̄ɤ/ 'young' + /zɪ̌/ 'juvenile' + /yø̌/ 'garden' ➡ [ǰɪ̄ɤzíyø̀] 'kindergarten'

#9 /sɤ̄/ 'hand' + /tɕɪ́?/ 'finger' + /kʰʌ́?/ 'nail' ➡ [sɤ̄tɕɪ́kʰʌ?] 'finger nail'

#10 /sɻ̄/ 'water' + /mɪ̌?/ 'honey' + /dɔ̌/ 'peach' ➡ [sɻ̄mɪ́dɔ̀] 'a kind of peach'

#11 /zə̌ŋ/ 'spirit' + /tɕɪ̂ŋ/ 'essence' + /bɪ̌ŋ/ 'sickness' ➡ [zə̀ŋtɕíŋbɪŋ] 'crazy'

#12 /mɔ̌/ 'horse' + /ɕ̌ī/ 'drama' + /dø̌/ 'group' ➡ [mɔ̀ɕ́ídø̀] 'circus'

#13 /lɔ̌/ 'old' + /zɪ̌/ 'former' + /pe̋/ 'generation' ➡ [lɔ̀zípè] 'sage'

#14 /ɦɔ̌ŋ/ 'red' + /ɕɪó?/ 'blood' + /dǯɪ̄ɤ/ 'ball' ➡ [ɦɔ̀ŋɕɪ́odǯɪ̀ɤ] 'red blood cell'

#15 /dǔ/ 'big' + /zʌ̌?/ 'gate' + /hɑ̄/ 'crab' ➡ [dùzʌ́hà] 'crabs from Big Gate'

#16 /ɕɪ́?/ 'snow' + /hwɔ̂/ 'flower' + /kɔ̂/ 'cream' ➡ [ɕɪ́hwɔ́kɔ̀] 'vanishing cream'

#17 /tʰsʌ́?/ 'to exit' + /kʰɤ̄/ 'mouth' + /tsə̄ŋ/ 'document' ➡ [tʰsʌ́kʰɤ́tsə̀ŋ] 'document for leaving a country'

#18 /hʌ́?/ 'black' + /dɤ̌/ 'head' + /fʌ́?/ 'hair' ➡ [hʌ́dɤ́fʌ?] 'black hair'

#19 /pʰó?/ 'to strike' + /kʰʌ́?/ 'competent' + /bɑ̌/ 'placard' ➡ [pʰókʰʌ́bà] 'cards (poker)'

#20 /pʌ́?/ 'hundred' + /jɪ̌?/ 'page' + /tɕɪ́?/ 'knot' ➡ [pʌ́jɪ́tɕɪ?] 'a kind of food made of soybeans'

#21 /lɔ̌?/ 'to record' + /jîŋ/ 'sound' + /tɕî/ 'machine' ➡ [lɔ̀jɪŋtɕ̄ī] 'tape recorder'

#22 /zʌ̌?/ 'sun' + /pə̄ŋ/ 'original' + /ɲɪ̌ŋ/ 'people' ➡ [zʌpə̀ŋɲ̄ɪŋ] 'Japanese'

#23 /lɔ̌?/ 'green' + /dɤ̌/ 'bean' + /tʰɔ̂ŋ/ 'soup' ➡ [lɔdɤ̀tʰɔ̄ŋ] 'green bean soup'

#24 /mə̌?/ 'eye' + /tɪ́?/ '(particle)' + /vʌ̌?/ 'object' [mətɪ̀vʌ̄?] 'target'

#25 /tʌ̌?/ 'white' + /mɔ̌?/ 'wood' + /ɲɪ̌/ 'ear' ➡ [bʌmɔɲ̄ī] 'white fungus'

TABLE III. List of trisyllabic compounds and first, second and third members of the compounds.

QUADRISYLLABIC COMPOUNDS

#1 /ɕîŋ/ 'new' + /v̌əŋ/ 'to hear' + /tɕ̄ī/ 'to record' + /tsê/ 'person' ➡ [ɕíŋvə̀ŋtɕ̄ítsɛ̀] 'news reporter'

#2 /pø̄/ 'half' + /jɭɔ/ 'night' + /sɛ̂/ 'three' + /kɑ̂/ 'time measurement' ➡ [pø̄jɭɔsɛ̄kɑ̀] 'pre-dawn'

#3 /jɭ̌ɤ/ 'oil' + /tʰɕîã/ 'behavior' + /wʌ̌?/ 'slippery' + /dɪɔ/ 'tune' ➡ [jɭɤtʰɕɪ́ɑ̃wʌ̄dɪɔ̀] 'frivolous'

#4 /tɕɪ́?/ 'to unite' + /hwə̂ŋ/ 'matrimony' + /tsə̄ŋ/ 'proof' + /sɪ̂/ 'book' ➡ [tɕɪ́hwə́ŋtsə̄ŋsɪ̀] 'marriage licence'

#5 /zʌ̌?/ 'factual' + /zɻ̌/ 'to be' + /dǯɪ̄ɤ/ 'then' + /zɻ̌/ 'to be' ➡ [zʌ̀zɻ́dǯɪ̄ɤzɻ̀] 'unpretentious'

TABLE IV. List of quadrisyllabic compounds and the members of the compounds.

(´ = HIGH; ¯ = MID; ` = LOW; and ˆ = ´`; ˇ = `´; ¯́ = ¯´; ¯̀ = `¯) ('↑' = tone raised; '↓' = tone lowered)

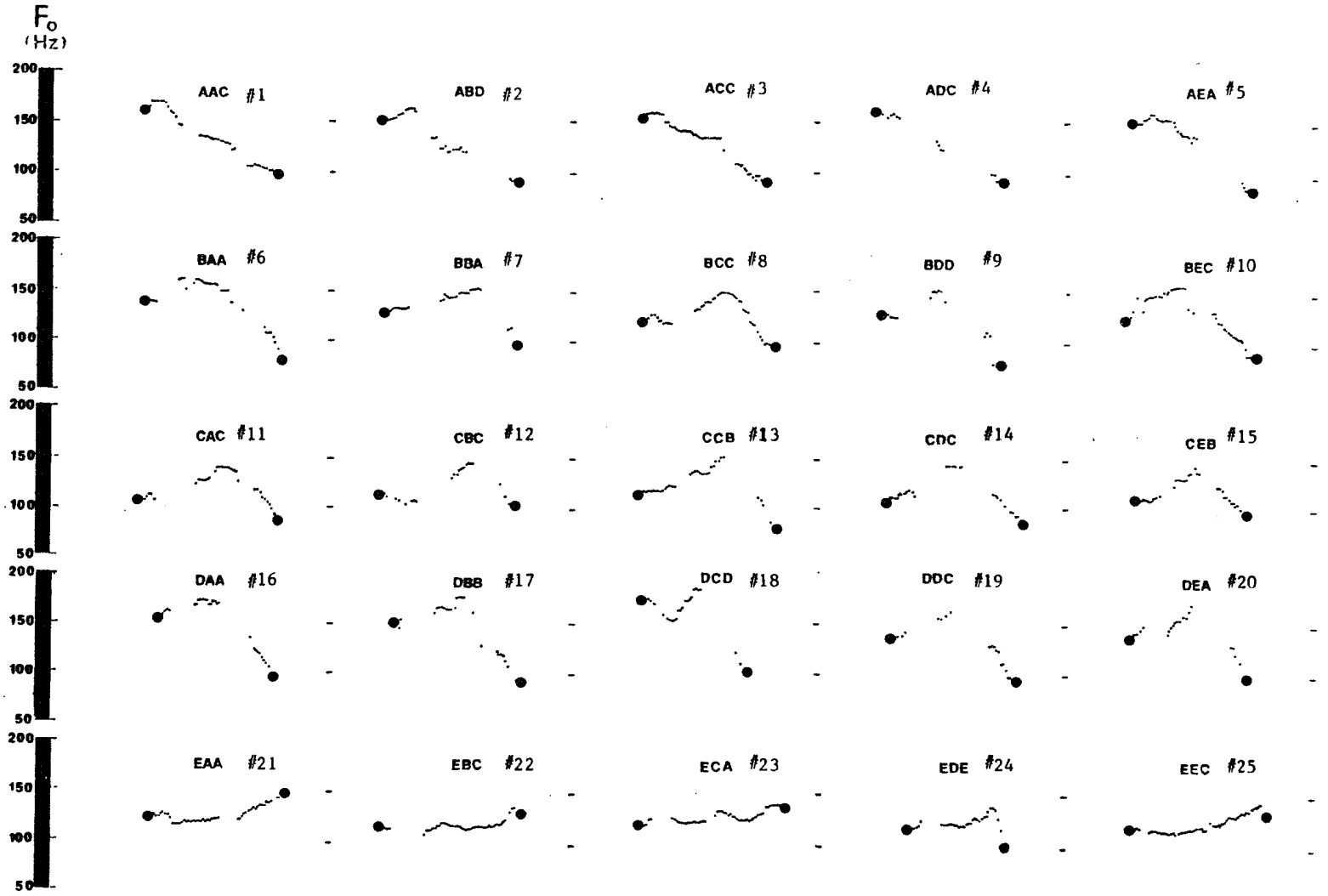


FIGURE 7. Fundamental frequency (pitch) contours of the trisyllabic compounds.

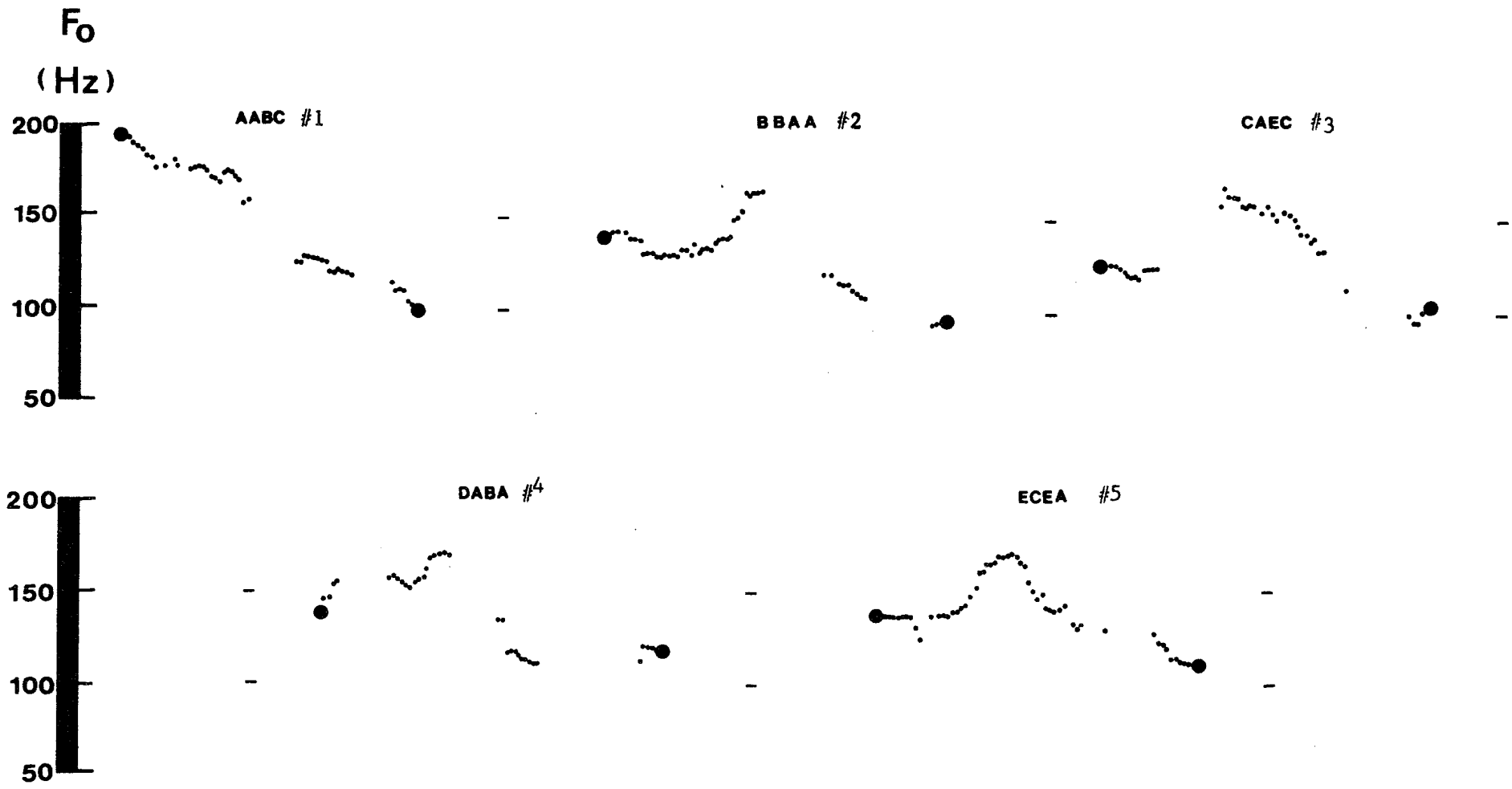


FIGURE 8. Fundamental frequency (pitch) contours of the quadrisyllabic compounds.

4. SUMMARY OF THE RESULTS OF THE ACOUSTICAL ANALYSIS.

1. The shapes of the five etymologically distinct Shanghai 'tones' that occur on monosyllabic words have been analyzed, confirming previous descriptions of the pitch patterns involved.

2. We find there are five main contours that occur on the bisyllabic compounds, plus an additional pattern in the cases where the second syllable is a type D syllable and the first is a type B, C or E syllable.

3. The shape of the pitch contour of a bisyllabic compound is similar to the shape of the pitch contour of its first syllable when it occurs as a monosyllabic word, except for the cases where the second syllable is a type D syllable and the first is a type B, C or E syllable.

4. There are five pitch contours that occur on the trisyllabic compounds.

5. There are only four pitch contours that occur on the quadrisyllabic compounds.

6. The shapes of the pitch contours of the bisyllabic, the trisyllabic and the quadrisyllabic compounds with a type A syllable as their first syllable are all similar to the pitch contour of a type A syllable in isolation.

7. The shapes of the pitch contours of the bisyllabic and the trisyllabic compounds with a type E syllable as their first syllable are similar to the pitch contour of the type E syllable in isolation (given that the final falling portion of the pitch contour of the type E syllable is not considered an intrinsic tonal property).

8. Although there is a difference in the first portion of the pitch contour, the remaining portion is similar in trisyllabic compounds with a type B, C or D syllable as their first syllable.

9. Although there is a difference in the first portion of the pitch contour, the remaining portion is similar in quadrisyllabic compounds with any type of syllable as their first syllable, namely, it falls to low.

10. The first portion of the pitch contours of the trisyllabic and the quadrisyllabic compounds is similar to the beginning points of the pitch contour of the first syllable in isolation.

5. PHONEMIC TONES.

We have discussed the pitch contours found for each of the syllable types A, B, C, D and E in Section 2. We will now present a phonological interpretation of Shanghai tones and tone sandhi.

In monosyllabic words only four contrastive pitch patterns are found. Type A syllables have a high falling pattern, type B a mid rising pattern, types C and E a low rising pattern and type D a high level pattern. Types C and E differ from each other in segmental composition; type C syllables being open or nasal final and long and type E syllables being checked by

a glottal stop and short. We may therefore say that syllables of types C and E do not differ in tone. Although type D syllables are also short and checked, their pitch is not similar to any of the long syllables A, B or C. We need therefore to provide for no more than four different contrastive tone patterns on monosyllabic words.

The question arises whether less than four tonemic contrasts should be recognised in Shanghai phonology. There are two aspects to consider, the relationship between consonants and tones and the possibility of decomposing the contours into more primitive elements. Because voiced obstruents are restricted to syllables of types C and E, it could be argued that the tone pattern of C and E (low rising) is a variant of the same phonological tone(s) as B (mid rising). There would then be a phonological rule lowering the tone after an initial voiced obstruent, and only three underlying tonal patterns on syllables would be recognised (A falling; B, C and E rising; D level). While this rule would provide a partial recapitulation of Shanghai historical tonology there are two principal arguments against its adoption. In the first place, the correlation of consonant voicing and tonal pattern breaks down when sonorants are examined: voiced nasals, laterals and approximants may appear initially in syllables of types A, B and D as well as C and E. Some examples of contrasting sets are given in (5):

(5)

A.	/wə̂ŋ/	"temperature"	A.	/wê/	"fearless'
B.	/wə̄ŋ/	"stable"	B.	/wē/	"to feed'
C.	/wə̌ŋ/	"faint"	C.	/wê/	"for"
A.	/mû/	"devil"	D.	/wʌ́?/	"to pick'
B.	/mū/	"female"	E.	/wʌ̌?/	"slippery"
C.	/mǔ/	"to grind"			

Secondly, the voicing of an obstruent is irrelevant in the process of tone sandhi: tones on non-initial syllables change in the same ways regardless of whether these syllables have voiced or voiceless initial obstruents, and no changes in obstruent voicing occur as a result of the tone changes[3]. We therefore maintain that four tonal patterns must be distinguished with types C and E distinguished from type B by some tonal property as well as by initial consonant voicing where necessary. Of course, the important redundancy involved in the relationship of obstruent voicing and tonal category needs to be stated somewhere in the grammar, but it is not a phonological rule which permits a reduction of the number of contrasting tonal patterns. In many tone languages, rising and falling pitch patterns can be shown to be due to the juxtaposition of a sequence of level tones. In other words, a falling contour may be analyzed as High followed by Low and a rising contour is analyzed as Low followed by High. In Shanghai, only type D syllables have a simple level (High) tone and no monosyllabic words with Mid or Low pitch occur. However, we believe there is evidence in favor of analyzing the pitch contours as being sequences of level tones.

3. Sherard (1972) reports that with a following Low tone the voiced stops have a slight breathy (murmured) quality. When the tone is changed the breathiness is absent but the voicing remains.

This evidence is principally to be found in the relationship between the pitch patterns on monosyllabic words and the pitch patterns on bisyllabic compounds and over the first portion of longer compounds. This relationship suggests that the contours should be decomposed and gives an indication of the nature of the underlying elements. As noted in Section 3, the pattern over the first two syllables of a compound is, with some exceptions, similar to that found when its first element occurs in isolation. For example, /di/ 'electricity' has a low rising pitch pattern in isolation; in a compound with /t^hi/ 'stairs', the first syllable has a low level pitch and the second syllable a mid level pitch (see #16 on Figure 6). What appears to have happened is that a sequence of 2 tone levels on the first syllable has been split so that each syllable receives one tone. The phonetic sequences of tones on monosyllables might be written in terms of 3 principal levels, High (H), Mid (M) and Low (L) as follows (where M↑ represents a raised Mid level):

SYLLABLE TYPE	PHONETIC TONE(S)
A	[HL]
B	[MM↑]
C, E	[LM↑]
D	[H]

Note however that in the case of the phonetic sequences [MM↑] and [LM↑] there is an alternation with what might be represented as [MH] and [LH] respectively. For example, in /di/ 'electricity' and in /di t^hi/ 'escalator' the contour is [LM↑], but in longer compounds with /di/ or another type C syllable as their first element the second syllable of the compound has a high level pitch (see the 3rd row of examples on Figure 7). The rule governing these alternations seems to be that lowering of High takes places when it follows a non-High and precedes a word-boundary. In other words the underlying tones for the monosyllables of the various syllable types are as follows:

SYLLABLE TYPE	PHONEMIC TONE(S)
A	/HL/
B	/MH/
C, E	/LH/
D	/H/

A more formal account of this analysis and further evidence in its favor will be presented below in the detailed statement of sandhi rules.

These rules will be formulated in terms of the revised framework for a suprasegmental analysis of tone proposed by Leben (1978). In an earlier proposal (Leben, 1971b, 1973a, b) Leben assumed that for a language with an underlying suprasegmental level of representation, a phonological rule would map the units in the suprasegmental tone patterns into segmental tone features, and that in phonetic representation tone was expressed as a segmental feature. In the revised version (Leben, 1978), the assumption that one is mapped into a segmental feature has been abandoned, instead Leben accepts Goldsmith's theory of 'autosegmental association between tones and segments" (Goldsmith, 1976a, b) which asserts that the units of suprasegmental tone patterns are not mapped into segmental tone features, instead the tonal representation is associated with the segmental representation at some stage in a phonological derivation but remains suprasegmental at all stages. Before the process of association, tones simply form a property of the word as a whole. After this process, association lines specify which segment(s) or syllable(s) each tone is coarticulated with. Derivations are subject to what Goldsmith has termed the "well-formedness condition":

(6) Well-formedness Condition (WFC):
  a. Every tone is associated with some syllable.
  b. Every syllable is associated with some tone.
  c. Association lines may not cross.

According to Goldsmith,

> 'the "well-formedness condition" is in the indicative not the imperative. A derivation containing a representation that violates the WFC is not thereby marked as ill-formed; rather, the Condition is interpreted so as to change the representation minimally by addition or deletion of association lines so as to meet the Condition maximally.' (Goldsmith, 1976b, p. 27).

This theoretical framework has been chosen because it seems to predict certain processes in Shanghai tone sandhi that would be unexpected if tone was a segmental property.[4] For example, it will be necessary to propose rules that delete tones without deleting the segment(s) with which they are lexically connected. Furthermore there is a process which places a low tone at the end of longer compounds; this tone may be realized over one, two or more syllables. Both of these cases seem to attest to the independence of the tonal and segmental levels in the phonological representation. This question will receive further comment in the course of the formal presentation of the rules.

6.0. TONE SANDHI PATTERNS.

We now turn to the phonological analysis of the tone sandhi patterns

[4] The second author is persuaded of the elegance and simplicity of the autosegmental account of the data but maintains reservations concerning the necessity of positing non-segmental elements in phonology.

observed on compounds. In order to do this, we will first establish a phonetic notation of the sandhi patterns, derived from the acoustical data in Figures 6, 7 and 8. As noted in Sections 3 and 4, the pitch contour of a bisyllabic compound is governed by the underlying tone(s) on the first syllable. In most cases, the pattern found on the first syllable when spoken in isolation is extended to include the second syllable. We therefore represent the patterns on the bisyllables with the same phonetic tones found on monosyllables. Note, however, that type C and type E syllables, although they have the same tones, result in slightly different patterns on bisyllabic compounds of which they form the first element. A type C intial syllable produces the pattern [L M↑], where a type E syllable produces the pattern [L LM↑].

When the second syllable is type D and the first is type B, C or E, the second syllable is higher pitched than when any other type of syllable folows B, C or E. The relationship between the phonetic tones in monosyllables and the bisyllabic compounds is shown in Table V:

	PHONETIC TONE(S) ON MONOSYLLABIC WORDS	PHONETIC TONE(S) ON BISYLLABIC COMPOUNDS WITH RESPECTIVE FIRST ELEMENTS		
				Type D as 2nd element
type A	[HL]	[H L]		
type B	[MM↑]	[M M↑]	~	[M H]
tone C	[LM↑]	[L M↑]	~	[L H]
tone D	[H]	[H H]		
tone E	[LM↑]	[L LM↑]	~	[L H]

TABLE V. Tone patterns on the bisyllabic compounds.

The pitch patterns on the longer compounds are similarly governed by the underlying tone(s) on the first syllable. The first two syllables receive the extended tone pattern of the first element of the compound and the remaining ones are generally low in pitch. A syllable between a high-pitched syllable and a low-pitched syllable will be mid in pitch. As with the bisyllabic compounds, an exceptional pattern occurs if a trisyllabic compound has a type E syllable as its first element. In this case the pattern over the three syllables will be [L L LM↑]. However, quadrisyllabic compounds with a first syllable of type E or type C have the same tone patterns as each other and are not exceptional. Note also that the first of two high-pitched syllables is lower than the second. Table VI presents the phonetic tones of these polysyllabic compounds in relation to the underlying and phonetic tones of the monosyllables occurring as first element.

FIRST ELEMENT OF COMPOUND	UNDERLYING TONES AS MONOSYLLABIC WORDS	PHONETIC TONES AS MONOSYLLABIC WORDS	PHONETIC TONES ON TRI-SYLLABIC COMPOUND WITH RESPECTIVE SYLLABLE TYPES AS FIRST ELEMENT	PHONEMIC TONES ON QUDRI-SYLLABIC COMPOUND WITH RESPECTIVE SYLLABLE TYPES AS FIRST ELEMENT
type A	/HL/	[HL]	[H M L]	[H M L↑ L]
type B	/MH/	[MM↑]	[M H L]	[M H M L]
type C	/LH/	[LM↑]	[L H L]	[L H M L]
type D	/H/	[H]	[H↓ H L]	[H↓ H M L]
type E	/LH/	[LM↑]	[L L LM↑]	[L H M L]

TABLE VI. Tone patterns on the trisyllabic and the quadrisyllabic compounds.

6.1. TONE RULES FOR THE DERIVATION OF THE TONE PATTERNS ON THE COMPOUNDS.

The phonetic tone patterns on the compounds can be derived by a number of simple rules which operate in conjuction with the Well-formedness Condition (WFC). The process of compound formation is viewed as one in which the word boundaries are deleted between the morphemes involved. This may be stated by the informal rule (7).

(7) WORD BOUNDARY DELETION.
When a compound is formed, the internal word boundaries are deleted.

As has already been established, the tone pattern of a compound is determined by the tone(s) on the first syllable only (with one exception). Because the tone(s) on any subsequent syllable make no contribution to the pattern we assume they have been deleted. It can be shown that they are not just subject to an assimilatory change under the influence of the tone(s) on the first syllable, since they are replaced on later syllables by an arbitrary inserted tone. Note that no other features are changed in the course of tone-deletion. Thus, for example,[a] type E syllable retains its initial voicing of an obstruent[4], its short vowel duration and its final glottal stop when it is in second position in a compound (the glottal stop tends to be lost when a third syllable follows).

We will argue that the deletion process does not apply when the second syllable of a bisyllabic compound is a type D syllable (i.e., has a single underlying tone, /H/) and the first syllable has /MH/ or /LH/ underlying tones. Hence the rule describing the process must be formulated with conditions to prevent its application in this environment. This rule is given as (8) where S stands for a syllable and T for the tone or tones associated with that syllable.

(8) TONE DELETION 1.

$$\begin{array}{ccccccccccc} \#S_1 & + & S_2 & + & \langle S_3 & + & \langle S_4 & + & \ldots\# & \longrightarrow & \\ \| & & | & & | & & | & & | & & \\ \#T_1 & & T_2 & & T_3\rangle & & T_4\rangle & & \ldots\# & & \end{array} \quad \begin{array}{ccccccccc} \#S_1 & + & S_2 & + & \langle S_3 & + & \langle S_4 & + & \ldots\# \\ \| & & | & & | & & | & & | \\ \#T_1 & & \emptyset & & \emptyset\rangle & & \emptyset\rangle & & \ldots\# \end{array}$$

Condition: does not apply if

$$T_1 = \begin{Bmatrix} MH \\ LH \end{Bmatrix} \quad \text{and} \quad T_2 = H \quad \text{and} \quad S_3 \ldots = \emptyset.$$

It was suggested in Section 5 that the two rising sequences of tones on monosyllables are underlyingly /MH/ and /LH/. The phonetic output in monosyllables and in most bisyllabic compounds of these tone sequences

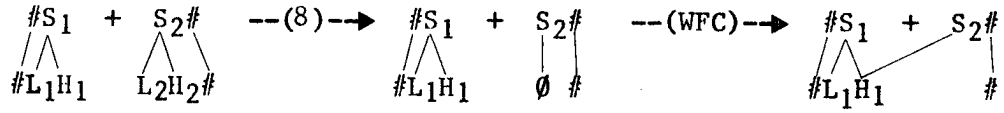
contains a lowered High tone, i.e., [MM$^{\downarrow}$] and [LM$^{\downarrow}$]. This argues for a tone lowering rule roughly of the form

(9) H-LOWERING.

$$H \longrightarrow M^{\downarrow} / \begin{Bmatrix} L \\ M \end{Bmatrix} \_\_ \#$$

Apparently the bisyllables with type D syllables in second position are exempt from this rule, as they retain a phonetic [H] tone on the second syllable. It seems explanatory to regard this exemption as due to the fact that these type D syllables retain their original underlying tone and do not receive their tone from the first syllable. Note that the Well-formedness Condition (WFC) will automatically re-associate a tone to the second syllable of bisyllabic compounds following TONE DELETION 1 (i.e., 8),for example:

(10)



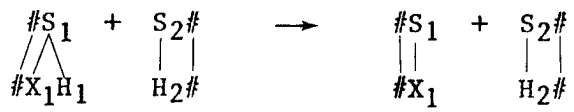
A language specific rule must apply to de-associate the second of two tones on the first syllable, so that only one tone is associated with each syllable. Using the subscript attached to the tones to impose a global condition on the tone-lowering rule,it can be reformulated as (9') to apply to just the monosyllables and those bisyllables which are affected;

(9') H-LOWERING (REVISED).

$$H_1 \longrightarrow M^{\downarrow} / \begin{Bmatrix} L \\ M \end{Bmatrix} \_\_ \#$$

This rule will not apply to bisyllables in which the original H tone remains on the second syllable, i.e., those not affected by TONE DELETION 1 (8). In these cases the tone sequence on the first syllable is simplified by deletion of the second tone. This process is described by the rule in (11):

(11) TONE DELETION 2.



In this rule 'X' is a cover symbol for any tone, although only M or L can occur in this position. Because the tone associated with S_2 is H_2, the HIGH LOWERING (REVISED)(9') does not apply to the output of TONE DELETION 2.

We have discussed in the earlier sections (3-5) how the spreading of the tones from the first syllable is essentially limited to affecting the second syllable of a compound. The tone on a third, fourth or later syllable is the same regardless of the initial syllable (with one exception). Although the tones on these syllables are deleted by TONE DELETION 1 (8), these syllables are not re-associated with the remaining tone(s) of the first syllable. A tone or tones must be inserted in the string instead. We will argue that this insertion rule simply associates a L tone with the third syllable of a compound. In the trisyllables the third syllable is phonetically Low, and in the quadrisyllables the final syllable is phonetically Low. There is evidence that a Low tone in the environment of H__L is raised to Mid. For example, a compound with a type A syllable as its first element (i.e., with underlying /HL/ on S_1) has the HL sequence spread over the first two syllables. When a third syllable occurs, the tone on the second syllable is raised to Mid. In a quadrisyllabic compound the tone on the third syllable is phonetically Mid when a phonetic High precedes and Low follows. We assume the same tone-raising rule has applied in this instance. Therefore the inserted tone on the third syllable can be taken to be invariably a Low tone. And as the tone on the fourth (and subsequent) syllable is Low on the surface it can be presumed that the Low tone inserted onto the third syllable is spread to any following syllables in the compound. In fact the Well-formedness (WFC) would automatically associate any following syllables (which are not associated with any tone following TONE DELETION 2 (11)) with the inserted tone. The insertion rule applies to all trisyllabic or longer compounds except trisyllabic compounds with a type E syllable their first element, so the rule can be formulated as (12):

(12) L-INSERTION.

$$\begin{array}{ccc} S_3 & \rightarrow & S_3 \\ | & & | \\ \emptyset & & L \end{array}$$

except if $S_4 = \emptyset$ and $S_1 = (C)V?$ (with V? = L H)

The exceptional trisyllabic compounds are similar to the bisyllabic compounds with an initial type E syllable in that both of the underlying tones on the first syllable are associated with the final syllable of the compound. This and other matters not automatically accounted for by the WFC are governed by a series of rules which regulate the association between tones and syllables. These rules are as follows:

(13) ASSOCIATION RULES.

(a) Delete the association line between the second of two tones and S_1 in any compound.
(b) Insert an association line between an initial L and the final syllable in a bisyllabic or trisyllabic compound, where S_1 = (C)V?.
(c) Insert an association line between S_2 and the tone to the left when $S_3 \neq \emptyset$ and S_1 = (C)V?. (H)

The operation of these association rules and the previously discussed tone deletion and insertion rules will be illustrated by showing the derivation of some selected compounds.

6.3. DERIVATION OF TONE PATTERNS ON BISYLLABIC COMPOUNDS.

The majority of bisyllabic compounds undergo TONE DELETION 1 (8), and ASSOCIATION RULE (13a). A simple derivation of this kind can be illustrated by a compound which has a type C syllable as its first element and a type A syllable as its second element. The monosyllables /di/ (LH) 'eletricity' and /t^hi/ (HL) ' stairs' combine to form a compound meaning 'elevator' (#11 in Figure 6 and Table 2). Following deletion of the word boundaries between these elements, the condition for TONE DELETION 1 (8) is satisfied, i.e.

(13a) #di + t^hi# --(8)--> #di + t^hi#
#L_1H_1 H_2L_2# #L_1H_1 Ø#

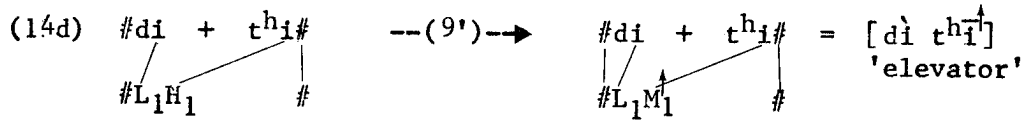
The immediate output of (8) violates the WFC because the second syllable is not associated with any tone, hence an association line is inserted as follows:

(14b) #di + t^hi# --(WFC)--> #di + t^hi#
#L_1H_1 Ø# #L_1H_1 #

The ASSOCIATION RULE (13a) serves to capture the generalization that in compounds of any length the first syllable may be associated with only one tone. It now applies in this derivation, viz.

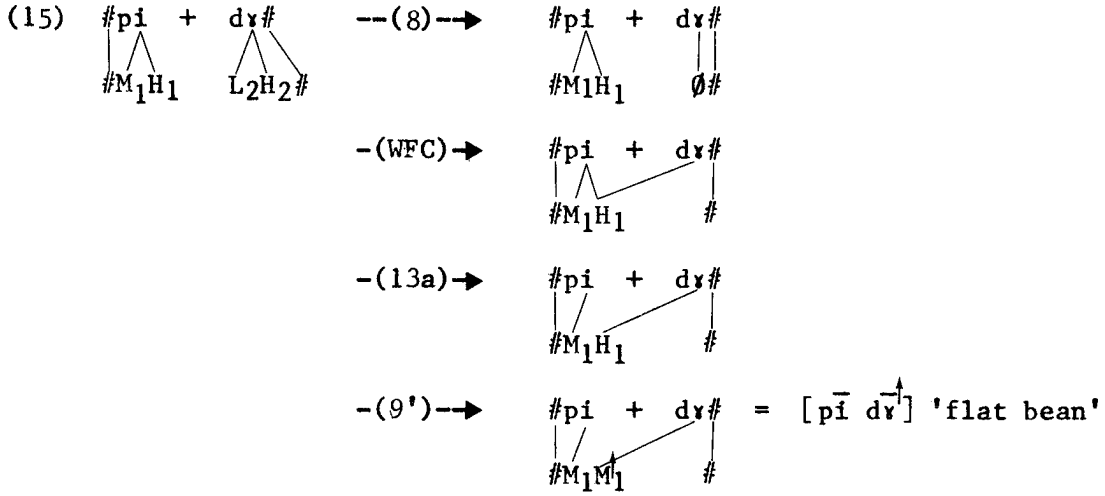
(14c) #di + t^hi# --(13a)--> #di + t^hi#
#L_1H_1 # #L_1H_1 #

The sequence $\#L_1H_1\#$ is subject to the H-LOWERING (REVISED)(9'), so that the eventual output from this derivation is as in (14d)

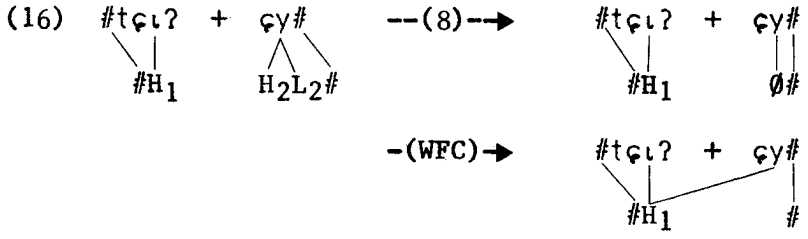
(14d) 

Note that there is no need to impose any extrinsic ordering of these rules. They apply whenever their structural description is satisfied.

An additional example of a compound which undergoes the same rules is one with a type B syllable as its first element and other than type D syllable as second element. The example is /pi/ (MH)'flat' + /dɤ/ (LH) 'bean' (#8 in Figure 6 and Table II) which has the derivation in (15).

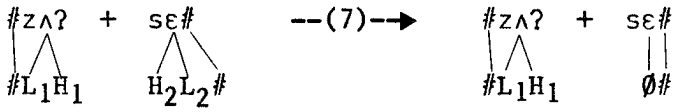
(15) 

When there is a single tone on the first element of a compound there is no work for rule 13a to do. The compound formed from /tɕɪʔ/ (H) 'urgent' and /ɕy/ (HL) 'need' (#16 in Figure 6 and Table II) has the derivational history shown in (16).

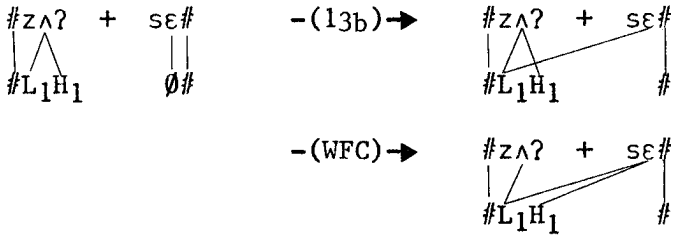
(16) 

The final output is [tɕɪ́ ɕý] 'urgent need' where the usual deletion of a medial glottal stop has also occurred.

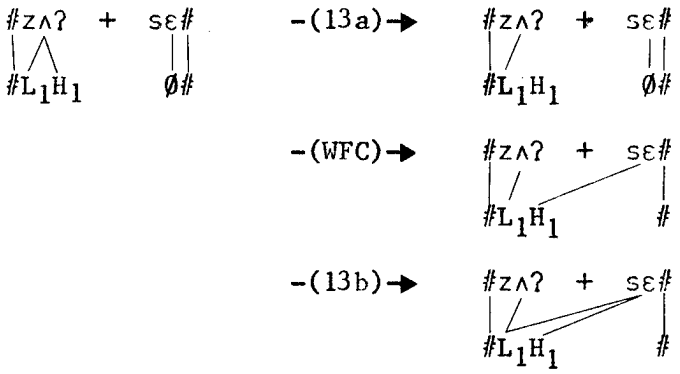
Where the first element of a bisyllabic compound is type E, TONE DELETION 1 (8) applies as with (14-16). However, ASSOCIATION RULE (13b) applies. An example is /zʌʔ/ (LH) 'ten' + /sɛ/ (HL) 'three' (#21 in Figure 6 and Table II). The tones on the 2nd syllable are deleted:

(17a) 

The ASSOCIATION RULE (13b) directs that the initial L be associated with the final syllable. When the required association line is inserted, the WFC intervenes to ensure that association lines do not cross. The minimal change which will satisfy the WFC is to associate the H tone to the final syllable and delete the association line between H and the first syllable. Thus,

(17b) 

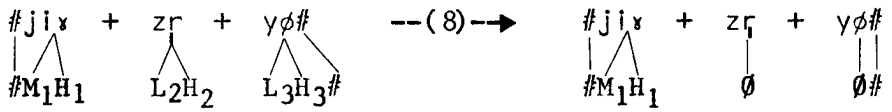
It is possible that (13a) applies before (13b) in this case. We would then have (17c) instead of (17b).

(17c) 

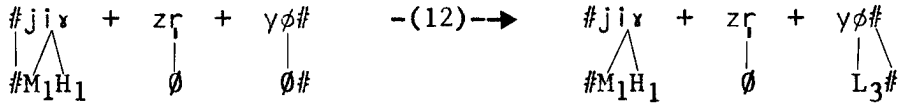
The output of these derivations is the same and there seems no need to choose between them. To do so would require abandoning the simple condition of rule application that rules apply whenever their structural description is met. In any order of application the rules will produce the final output [zʌ̀ sɛ̄] 'thirteen' via the H-LOWERING (REVISED)(9') and the optional glottal stop deletion rule.

6.4. DERIVATION OF TONE PATTERNS ON LONGER COMPOUNDS.

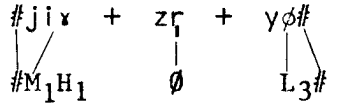
The longer compounds are affected by the L-INSERTION rule (12) as well as WORD BOUNDARY DELETION (7), the TONE DELETION 1 and 2 (8, 10) and the ASSOCIATION RULES (13). The most frequent kind of derivation can be exemplified by a trisyllabic compound which has a type B syllable as its first element, for example a compound formed from /jiɤ/ (MH) 'young' + /zɽ/ (LH) 'juvenile' + /yɸ/ (LH) 'garden' (#8 in Figure 7 and Table III). Following deletion of the internal word boundaries, TONE DELETION 1 (8) applies:

(18a) 

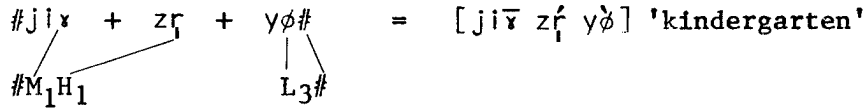
As there is a third syllable in this compound a Low tone is inserted into the tone string by L-INSERTION (12):

(18b) 

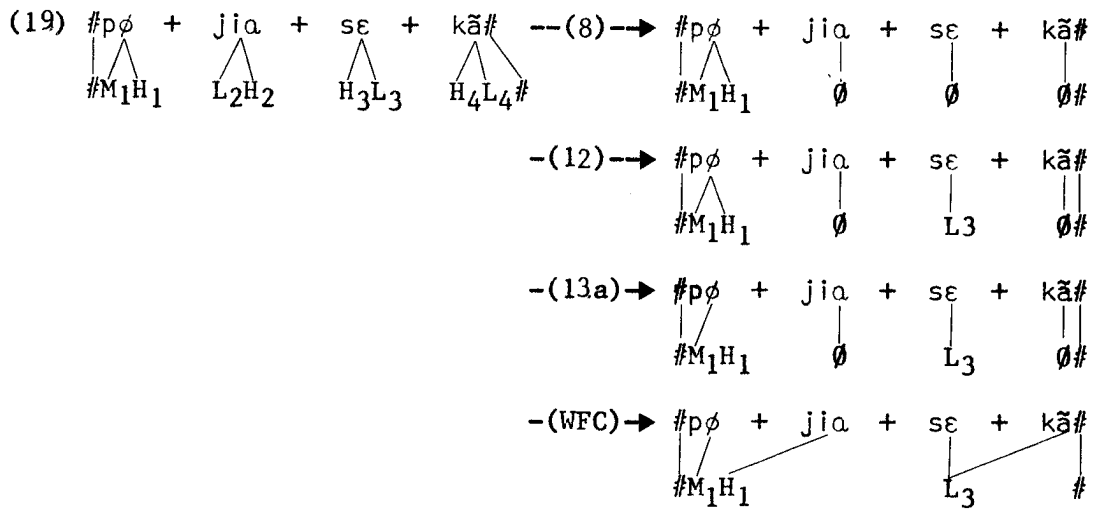
Of course, the output of (12) in (18b) as represented here violates the WFC since the second syllable is not associated with any tone. However, ASSOCIATION RULE (13a) directs that the association line between S_1 and the second of two tones be deleted. This produces the output in (18c):

(18c) 

The WFC intervenes to correct the violation by making the minimal change, that is, an association line between the unassociated second tone and the second syllable is inserted. The output is (18d) with tone pattern [M H L].

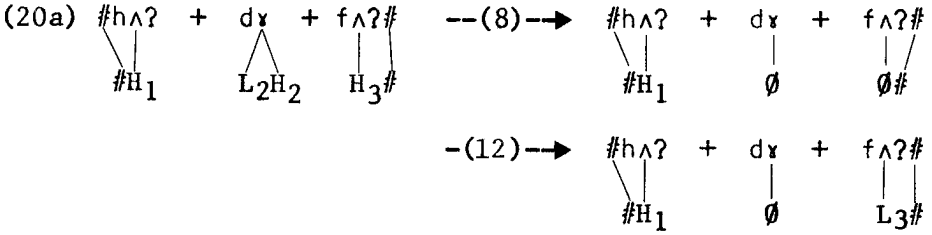
(18d) 

The derivation of a quadrisyllabic compound with an initial syllable of the same type is essentially similar, except that the WFC associates the inserted Low tone on S_3 to the fourth syllable as well. An example is the compound formed by combining /pɸ/ (MH) 'half' + /jiɑ/ (LH) 'night' + /sɛ/ (HL) 'three) + /kã/ (HL) 'time measurement' (#2 in Figure 8 and Table IV) whose derivation is given in (19).

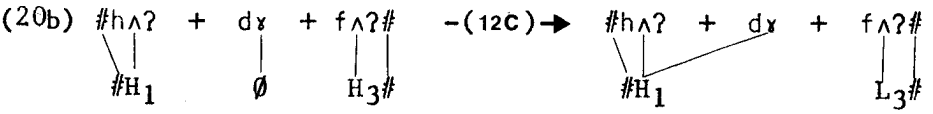


The eventual phonetic output from this derivation shows the effect of the tones on each other. Because of co-articulation effects between H on the second syllable to the following L, the pitch of the third syllable is closer to a phonetic Mid level and the compound is transcribed [pø̄ jiá sɛ̄ kã] 'pre-dawn'. As this kind of coarticulation between adjacent tones is a straightforward phonetic effect it is not discussed further in this presentation and no rule has been formulated.

In the derivation of longer compounds with an initial type D syllable the ASSOCIATION RULE (13c) applies. For example, the trisyllabic compound formed from /hʌ?/ (H) 'black' + /dɤ/ (LH) 'head' + /fʌ?/ (H) 'hair' (#18 in Figure 7 and Table III) has the derivation in (20).



The output of (12) violates the WFC because the second syllable is not associated with any tone. However, the WFC alone does not uniquely resolve this case since S_2 might be associated with either the tone to the left or the tone to the right. The correct output must associate the second syllable with the H tone on S_1 and (13c) inserts an association line which does so.



There seems to be a slight lowering of the pitch of the first of two syllables associated with an initial H tone when a third syllable follows in the compound. This lowered high pitch is represented in Table IV by H↓ and the final phonetic output from this derivation is [hʌ́↓ dɤ́ fʌ̀?] 'black hair'. This phonetic lowering rule applies also to the quadrisyllabic compounds with an initial type D syllable, see #4 in Figure 8 and Table IV.

The longer compounds with an initial type E syllable undergo different derivations according to the number of syllables they contain. Trisyllables are affected by rule (13b), as are bisyllables with an initial type E syllable, but quadrisyllabic compounds share the same kind of derivation as most of other longer compounds exemplified in (19). A trisyllabic compound of this type is formed from /lo?/ (LH) 'to record' + /jɪŋ/ (HL) 'sound' + /tɕi/ (HL) 'machine' (#21 in Figure 7 and Table III) and its derivation is given in (21). Note that L-INSERTION (12) does not apply in this derivation.

(21) #lo? + jɪŋ + tɕi# --(8)→ #lo? + jɪŋ + tɕi#
#L$_1$H$_1$ H$_2$L$_2$ H$_3$L$_3$# #L$_1$H$_1$ Ø Ø#

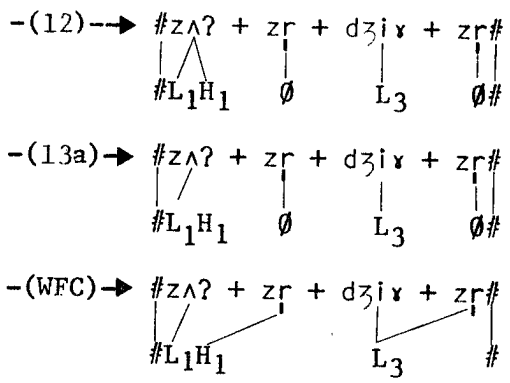
-(13b)→ #lo? + jɪŋ + tɕi#
#L$_1$H$_1$ Ø #

-(WFC)→ #lo? + jɪŋ + tɕi#
#L$_1$H$_1$ #

As with the derivation of the bisyllabic compound in (17) the WFC intervenes to correct the violation, namely crossed association lines, that would otherwise result from the application of (13b). Again as with (17), it is immaterial if (13a) does nor does not apply to this case. In either case the eventual output, including the effect of H-LOWERING (REVISED)(9'), is [lò jɪ̀ŋ tcT↓] 'tape recorder'.

The difference in the derivation of a quadrisyllabic compound which has a type E initial syllable can be illustrated with the compound formed from /zʌ?/ (LH) 'factual' + /zɾ/ (LH) 'to be' + /dʒiɤ/ (LH) 'then' + /zɾ/ (LH) 'to be'. Compare the derivation in (22) with that in (19). The derivation in (22) can be seen to be like the normal quadrisyllabic compound illustrated in (19).

(22) #zʌ? + zɾ + dʒiɤ + zɾ# --(8)→ #zʌ? + zɾ + dʒiɤ + zɾ#
#L$_1$H$_1$ L$_2$H$_2$ L$_3$H$_3$ L$_4$H$_4$# #L$_1$H$_1$ Ø Ø Ø#



The eventual phonetic output is [zʌ̀ zɾ́ dʒiɤ̄ zɾ̀] 'unpretentious' where the co-articulation between adjacent tones is noted by representing the pitch of the third syllable as in the mid range. Any other compounds undergo similar derivations.

7. CONCLUSION.

We now return to the question of the accuracy of previous statements in the linguistic literature on Shanghai tone sandhi. The regular patterns of sandhi we have found show two major processes, the spreading of the tone(s) of the first syllable of a compound to the second syllable, and the replacement of the tone(s) on third and subsequent syllables by Low tone. There are a few minor patterns in addition, but, in the majority of patterns, only the tone of the first syllable matters in the output.

Sherard's observations on tone sandhi do not agree entirely with ours. A significant difference between the data in this paper and the contours reported by Sherard concerns the longer compounds. Sherard reports that for compounds with type B or D initial syllables the contour for the compounds is same as that on the initial syllables when spoken as monosyllables (i.e., mid-dipping and high level, respectively). Sherard's data on compounds with type, C or E syllables agree in the main with ours. It seems to us that the differences arise because of two factors. Principally, as Sherard himself notes, the linguistic situation in Shanghai has been much influenced by immigration from surrounding areas. In particular, the speech variety that Sherard records seems to contain features characteristic of the dialect of Ningpo[5], to the south of Shanghai. In Ningpo, disyllabic compounds with type B and D initials do have the contours that Sherard reports. No Low tone insertion applies to these cases. For example, Sherard cites the compound formed from /ʌʔ/ (H) 'to suppress' + /se/ (MH) 'age' + /di/ (LH) 'money', referring to money given at the New Year, as [ʌ́ sé dí] with a High tone throughout, rather than as [ʌ́ sé dì] with final Low, which is the form that would occur in our sub-

[5] Ningpo is a town about 100 miles south of Shanghai. The first author is personally acquainted with the speech of this area through friends and relations who originate from there.

ject's speech. In some other cases it appears that Sherard may be reporting a form that was not read as a true compound. Our overall conclusion is that our data cannot necessarily be interpreted as correcting Sherard where they differ. Our data is more complete than Sherard's but apparently refers to a slightly different dialect. We would claim that our subject's speech is more truly representative of the indigenous Shanghai dialect whose historical connections have been more with the other Northern Wu dialects surrounding the city until recent times.

Ballard (1977) claimed that tone sandhi in Shanghai consists only of a rightward spreading (or in our terms a rightward reassociation) of the tone(s) of the initial syllable. This claim is more justified on the basis of the data in Sherard than it would be for ours but even so it does not account for all the patterns that Sherard reports. For example a compound beginning with a type C syllable (with a low-mid rise in isolation) has a pattern which "starts from a low level, rises to a moderately high pitch and then begins to fall back toward low level" (Sherard, 1972: 102). We recognise our High-Lowering (9') and Low-Insertion (12) rules in relating the tones on the monosyllable and the compound here. Such examples show that there is more than rightward spreading involved even in the sub-dialect Sherard examined. In particular, there is no case for regarding the sandhi process as one in which a contour is extended as a unit over the compound. With the majority of syllables in Shanghai having two underlying tones, it is natural that it is precisely the third syllable of a compound which receives an inserted Low tone, and while this single Low tone may spread over (be associated with) several syllables, contours as such do not spread.

REFERENCES

Anderson, S. R. (1978) "Tone features" Tone: A Linguistic Survey (V. A. Fromkin, ed.) New York, Academic Press: 133-175.

Ballard, W. L. (1977) On some Aspects of Tone Sandhi. Atlanta, Georgia State University.

Goldsmith, J. (1976a) "An overview of autosegmental phonology." Linguistic Analysis, 2, 23-68.

Goldsmith, J. (1976b) Autosegmental Phonology. Ph.D. Dissertation (MIT). Repr. by Indiana University Linguistics Club.

Hangyu Fangyan Gaiyao (A precis of Chinese dialects, 1960). Peking, Wenzi Gaige Chubanshe.

Hombert, J-M (1978) "Consonant types, vowel quality, and tone" Tone: A Linguistic Survey (V. A. Fromkin, ed.) New York, Academic Press: 77-111.

Jiangsushen He Shanghaishi Fangyan Gaikuang (Synopsis of Shanghai dialect and the dialects in Jiangsu Province, 1960). Nanking, Jiangsu Renmin Chubanshe.

Kennedy, G. A. (1953) "Two tone patterns in Tangsic" Language 29: 367-373.

Leben, W. R. (1971) "Suprasegmental and segmental representation of tone." Studies in African Linguistics, Supp. 2, 183-200.

Leben, W. R. (1973a) "The role of tone in segmental phonology." Consonant Types and tone (L. Hyman, ed.). Los Angeles, University of Southern California, Linguistics Program: 115-149.

Leben, W. R. (1973b) Suprasegmental Phonology. Ph.D. Dissertation (MIT).

Leben, W.R. (1978) "The representation of tone" Tone: A Linguistic Survey (V. A. Fromkin, ed.). New York, Academic Press: 177-219.

Ohala, J and Ewan, W. (1972) "Speed of pitch change." Paper presented at the 94th ASA Meeting, Autumn, 1972, Miami Beach, Florida.

Sokolov, M. V. (1965) "An experimental investigation of the tones in the Shanghai dialect." Phonetica, XII, 197-200.

Sherard, M. (1972) Shanghai Phonology. Ph.D. Dissertation (Cornell University).

Traill, A. (in progress) Phonetic and Phonological Sketches of !xóõ Bushman. Ph.D. Dissertation (University of Witwatersrand).

Effect of vowel quality on perception of nasals in noise

Eric Zee

[Paper presented at the Joint Meeting of the Acoustical Society of America (96th Meeting) and the Acoustical Society of Japan, Nov. 27-Dec. 1, 1978]

1. INTRODUCTION

There have been a number of studies on the perception of nasal consonants. The results of these experiments confirm that formant transition plays an important role in identification of nasal consonants, whether the formant transition follows or precedes the nasal consonants and whether the stimuli for the experiments are synthetic or natural speech sounds. Liberman, Delattre, Cooper and Gerstman (1954), using synthetic VC stimuli, investigated the role of vowel-consonant transition in the perception of syllable final stops and nasals, m, n, ŋ. In this experiment, a fixed three formant nasal segment was used for all nasal consonants. They showed that a variable F2 transition of two formant vowels alone could serve as a cue for the identification of the place of articulation of these obstruents. Nakata (1959) also demonstrated that F2 transitions were important perceptual cues for the identification of nasal consonants, m, n, ŋ, in synthetic CV stimuli. Malécot (1956) conducted a similar investigation by using real speech stimuli which contained unaltered, spliced and edited CV and VC syllables. In this experiment, æ was the only vocalic element in the stimuli. The results again indicated that the transitions of the adjoining vowels were important cues in the perception of m, n, ŋ. Furthermore, nasal resonances were found to serve primarily as class markers, differentiating nasal from stop consonants: they were not, however, completely neutral with respect to the identification of place of articulation. Nord (1976) investigated the relative importance of nasal murmur compared to vowel transitions as a cue for the identification of nasal consonants, by exchanging the nasal segment or the nasal segment together with a part of the CV and/or VC transition between nonsense words ama and ana uttered by Swedish speakers. He concluded that the nasal resonance contains some information on the place of articulation of the nasal consonants, although it was impossible to quantify the relation between the nasal murmur and the formant transition in this respect. Importantly, he found that more place information was conveyed by the transition into the next vowel than by the transition from the preceding vowel. Wang and Fillmore (1961), commenting on the importance of transitions, suggested that in a CVC syllable where C was p, t, k, b, d, g, m, n or ŋ and V was i, ɛ, a, ɔ or u, the consonant-vowel transition provided a stronger perceptual cue when the intrinsic amplitude of the vowel was greater.

The present study investigates the effect of vowel quality on the perception of nasal consonants in VC syllables in noise. V is i, e, a, o or u and C is m, n or ŋ. The VC syllables were masked with white noise of different levels of intensity. As different vowels have different intrinsic intensity, different intrinsic pitch and different formant structures, and as the patterns of formant transitions in

VC syllables containing different vowels are also different, we should reasonably expect that the identifiability of the syllable final nasals should vary according to the type of vowel that precedes the consonants. We should also expect different results with different nasal consonants. The purpose of the experiment is to investigate how different vowels condition differently the perception of nasal consonants in VC syllables in noise.

2. EXPERIMENT.

2.1. METHOD.

2.1.1. SUBJECTS.

The subjects were eight professors of phonetics and phonetically trained graduate students. There were five males and three females.

2.1.2. STIMULI.

A word list containing 60 tokens of nonsense VC syllables (5 vowels i, e, a, u, o x 3 nasal consonants m, n, ŋ x 4 repetitions of each VC combination) was prepared. The tokens were arranged in a random order. Two male phoneticians, both speaking standard American English, read the word list at a slow rate of careful speech. There was an approximately two second pause between tokens. The speakers were instructed to keep the pitch of the syllables level and were also asked to minimize any audible release of the nasal consonants. Out of 4 tokens, or repetitions, for each VC syllable type, one was chosen to become one of the 15 stimuli for the subsequent listening test on the basis that its pitch was most level and its intensity was within the range of 5 db compared to the other chosen tokens for the different VC syllable types. The 15 different VC syllable types for each speaker are listed as follows:

im	em	am	um	om
in	en	an	un	on
iŋ	eŋ	aŋ	uŋ	oŋ

The selected tokens of these syllables were rearranged in 4 quasi-random orders such that syllables having the same vowel never occurred as consecutive stimuli. This was done in order to minimize acoustic contrast effects on phonetic identification of the nasal consonants. These stimulus sets containing 60 syllables (4 x 15) were repeated

3 times. Each set was then masked by white noise which had a frequency range of 60-4000 Hz. Figure 1 shows a schematic block diagram of apparatus used for generating the white noise and for mixing the speech signals with the white noise. The white noise was generated by a Random Noise Generator (General Radio Co.) and was filtered by a Krohn-Hite Band Pass Filter, Model 330N. A recording of the white noise at three levels of intensity, 0 db, -6 db and -12 db, was made. The speech signal recorded earlier was then masked by the white noise by way of SOS (sound-on-sound) mixing. The db meter was a Brüel & Kjær Electronic Voltmeter, Type 2409 and the attenuator was a Decade Attenuator, Type 1450-TBR (General Radio Co.).

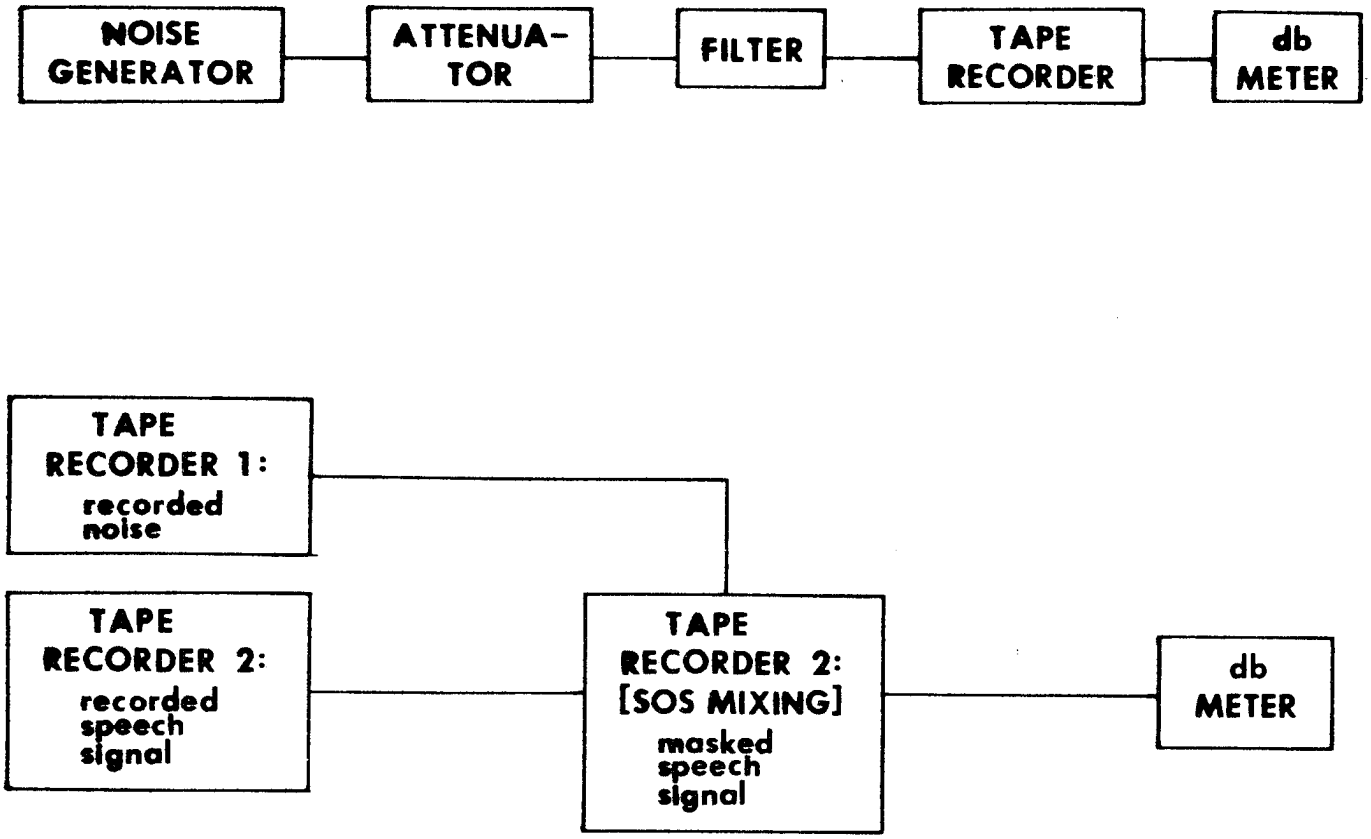


Figure 1. Schematic block diagram of apparatus used for the preparation of stimulus tapes.

The noise-to-speech ratios were in decreasing order, 16 db, -12 db, -18 db, for the three sets of recorded speech signals. A 0 db noise-to-speech ratio would mean that the intensity of the noise and the mid point intensity value of the speech signals are the same. Thus, there were three listening conditions and the stimulus tape for each of the two speakers contains 180 stimuli (3 noises x 15 syllable types x 4 quasi-random orders). There was a pause of approximately 2 seconds between stimuli and a longer pause of approximately 5 seconds after every 15 stimuli. The noise was, however, continuous throughout every 15 tokens, and broke off only at the end of every 15 tokens. The noise came on approximately one second before the first token started and stopped approximately one second after the last of the 15 tokens ended.

2.2. PROCEDURE.

The stimulus tapes were played to each of the 8 subjects through a high quality loudspeaker in a sound treated room. The subjects were professors and graduate linguistics students with phonetic training and whose native language was English. There were two separate sessions for each of the subjects, who listened to the stimulus tape of one of the speakers on one day and to the tape of the second speaker on a different day. Prior to each session, the subjects were given 15 stimuli which were also masked (noise-to-speech ratio: -6 db) for practice. The loudness of the stimuli was adjusted to each individual subject's preference. The subjects were asked to identify the syllable final nasal consonant in each VC syllable as m, n or ŋ. They were instructed to make a guess if they could not decide.

3. RESULT.

Table Ia and Table Ib are the two sets of responses of each of the 8 subjects to the stimuli containing the speech signals produced by the two speakers, G.P. and L.G. In each set, there are three groups of responses according to the ratio of noise to speech signal. The arrows under the intended nasal consonants in each vowel category mean 'perceived as'. Thus, in Table Ia, for instance, the upper left-most group, '1m and 3n', in [i] category, for Subject S.D., means that [im] was 1 time perceived as [im] and 3 times as [in] when the noise-to-speech ratio was -6 db. The total number of responses of all the eight subjects to each VC syllable type according to the three noise conditions are listed in Table IIa and Table IIb for the speech of Speaker G.P. and Speaker L.G. There is a total of 96 cases, or responses (3 noise conditions x 4 repetitions x 8 subjects) for each VC syllable type. Table III shows the combined number of the responses of the eight subjects to each stimulus type for the speech of both speakers. Again, the arrows here mean 'perceived as'.

SPEECH SIGNAL BY G.P.
RATIO OF NOISE LEVEL TO SPEECH SIGNAL: -6 db

STIMULI:	i	e	a	u	o
SUBJECTS	-m -n -ŋ	-m -n -ŋ	-m -n -ŋ	-m -n -ŋ	-m -n -ŋ
1. S.D.	1m 3n 4n 1n 3ŋ	4n 3n 1n 1ŋ 3ŋ	2m 1n 4n 1ŋ 4ŋ	3m 1n 4n 1n 3ŋ	1m 1n 3n 1n 3ŋ
2. J.W.	4n 4n 4n	4n 4ŋ 4ŋ	4m 4n 4ŋ	1m 2m 2n 2n 1ŋ 4ŋ	4m 3n 1n 1ŋ 3ŋ
3. P.L.	4n 4n 3n 1ŋ	1m 3n 3n 1n 1ŋ 3ŋ	4m 4n 4ŋ	2m 1m 1m 2n 3n 2n 1ŋ	4m 1m 3n 1n 1ŋ 2ŋ
4. S.G.	4n 3n 4n 1ŋ	4n 4n 4ŋ	4m 4n 4ŋ	3m 3m 3m 1n 1n 1n	4m 1m 3n 4n
5. R.J.	4n 4n 4n	1m 3n 4n 1n 3ŋ	1m 3n 4n 4ŋ	2m 1m 1n 3n 1ŋ 4ŋ	1m 1m 1m 2n 1n 3ŋ 1ŋ 2ŋ
6. I.M.	4n 4n 4n	4n 3n 1ŋ 4ŋ	2m 2n 4n 4ŋ	2m 2n 4n 3n 1ŋ	1n 4n 3ŋ 4ŋ
7. V.P.	4n 4n 4n	4n 3n 2n 1ŋ 2ŋ	4m 4n 4ŋ	2m 1m 4n 1n 1ŋ	2m 1n 4n 3n 1ŋ 1n
8. A.W.	1m 3n 4n 4n	4n 4n 2n 2ŋ	4m 4n 4ŋ	3m 2m 2n 1n 1ŋ 3ŋ	3m 1n 1n 1ŋ 3ŋ 3ŋ

SPEECH SIGNAL BY G.P.
RATIO OF NOISE LEVEL TO SPEECH SIGNAL: -12 db

STIMULI:	i	e	a	u	o
SUBJECTS	-m -n -ŋ	-m -n -ŋ	-m -n -ŋ	-m -n -ŋ	-m -n -ŋ
1. S.D.	2m 1n 4n 1n 1ŋ 3ŋ	1m 3n 3n 1ŋ 4ŋ	4m 4n 4ŋ	4m 4n 4ŋ	4m 4n 3n 1ŋ
2. J.W.	4n 4n 2n 2ŋ	4n 4n 4ŋ	4m 4n 4ŋ	4m 1m 3n 4ŋ	4m 4n 4ŋ
3. P.L.	4n 4n 3n 1ŋ	4n 4n 1n 3ŋ	4m 4n 4ŋ	4m 1m 4n 2n 1ŋ	4m 1m 4n 1n 2ŋ
4. S.G.	3m 1n 4n 4n	3m 1n 4n 4ŋ	4m 4n 4ŋ	4m 2m 2n 2n 2ŋ	4m 4n 4n
5. R.J.	1m 3n 3n 1n 1ŋ 3ŋ	4n 4n 2n 2ŋ	2m 2n 4n 4ŋ	4m 2m 1n 1ŋ 4ŋ	1m 4n 1n 3ŋ 3ŋ
6. I.M.	4n 4n 4n	4n 4n 4ŋ	2m 2n 4n 4ŋ	3m 1n 4n 2n 2ŋ	2m 1n 4n 1ŋ 4ŋ
7. V.P.	4n 4n 1n 3ŋ	1m 3n 4n 4ŋ	3m 1n 4n 4ŋ	4m 1m 3n 3n 1ŋ	2m 2n 4n 4n
8. A.W.	3m 1n 4n 4ŋ	4m 1m 3n 2n 2ŋ	4m 4n 4ŋ	4m 4n 4ŋ	4m 4n 4ŋ

SPEECH SIGNAL BY G.P.
RATIO OF NOISE LEVEL TO SPEECH SIGNAL: -18 db

STIMULI:	i	e	a	u	o
SUBJECTS	-m -n -ŋ	-m -n -ŋ	-m -n -ŋ	-m -n -ŋ	-m -n -ŋ
1. S.D.	3m 1n 4n 4ŋ	2m 2n 4n 4ŋ	4m 4n 4ŋ	4m 4n 1n 3ŋ	4m 1m 1m 3n 2n 1ŋ
2. J.W.	4n 4n 2n 2ŋ	4n 3n 1ŋ 4ŋ	4m 4n 4ŋ	4m 4n 4ŋ	4m 4n 4ŋ
3. P.L.	4n 4n 4n	4n 4n 1n 3ŋ	4m 4n 4ŋ	3m 1n 4n 1n 3ŋ	4m 4n 2n 2ŋ
4. S.G.	1m 3n 4n 2n 2ŋ	1m 3n 4n 4ŋ	4m 4n 4ŋ	4m 1m 4n 3n	4m 4n 4n
5. R.J.	4n 3n 1ŋ 4ŋ	4n 4n 4ŋ	3m 1n 4n 4ŋ	4m 1m 4n 3ŋ	4m 4n 2n 2ŋ
6. I.M.	1m 3n 4n 4ŋ	4n 4n 1n 3ŋ	3m 1n 4n 4ŋ	2m 2n 4n 4ŋ	1m 3n 4n 4ŋ
7. V.P.	4n 4n 4ŋ	4n 4n 4ŋ	3m 1n 4n 4ŋ	3m 1n 4n 4ŋ	2m 2n 4n 2n 2ŋ
8. A.W.	3m 1n 4n 4ŋ	4m 3n 1ŋ 4ŋ	4m 4n 4ŋ	4m 4n 4ŋ	4m 4n 4ŋ

Table 1a. Responses by individual subject to stimuli by Speaker G.P.

SPEECH SIGNAL BY L.G.
RATIO OF NOISE LEVEL TO SPEECH SIGNAL: -6 db

STIMULI:	i	e	a	u	o
SUBJECTS	-m -n -ŋ	-m -n -ŋ	-m -n -ŋ	-m -n -ŋ	-m -n -ŋ
1. S.D.	1m 3n 3n 3n 1ŋ 1ŋ	3n 4n 1ŋ 4ŋ	4m 4n 4ŋ	2m 3m 1n 2ŋ 4ŋ	4m 2n 2ŋ 4ŋ
2. J.W.	1m 1m 1n 2n 2n 2ŋ 2ŋ 1ŋ	2n 2ŋ 4ŋ 4ŋ	4m 3n 3n 1ŋ 1ŋ	4m 3m 1n 4ŋ	4m 2n 2ŋ 4ŋ
3. P.L.	3m 1n 4n 4n	1m 1n 2n 2n 2ŋ 2ŋ 2ŋ	4m 4n 4ŋ	1m 4n 3n 4n	4m 4n 4ŋ
4. S.G.	4n 2n 1n 2ŋ 3ŋ	4n 1n 3ŋ 4ŋ	4m 4n 4ŋ	4m 1m 3n 4ŋ	4m 1m 4n 2n 1ŋ
5. R.J.	1m 3n 1n 2n 3ŋ 2ŋ	4n 4n 1n 3ŋ	4m 4n 4ŋ	1m 1m 2n 3n 1ŋ 4ŋ	3m 4m 3n 1ŋ 1ŋ
6. I.M.	4n 3n 3n 1ŋ 1ŋ	4n 4n 4ŋ	4m 4n 4ŋ	2m 1n 4n 1ŋ 4ŋ	4m 4n 4ŋ
7. V.P.	4n 2n 1n 2ŋ 3ŋ	4n 3n 1ŋ 4ŋ	4m 4n 4ŋ	3m 2m 1n 1n 3n 1ŋ 1ŋ	4m 4n 4ŋ
8. A.W.	4n 4ŋ 4ŋ	1n 4n 4ŋ 3ŋ 4ŋ	4m 1m 4n 3ŋ	3m 3m 1ŋ 1n 4ŋ	4m 1n 3ŋ 4ŋ

SPEECH SIGNAL BY L.G.
RATIO OF NOISE LEVEL TO SPEECH SIGNAL: -12 db

STIMULI:	i	e	a	u	o
SUBJECTS	-m -n -ŋ	-m -n -ŋ	-m -n -ŋ	-m -n -ŋ	-m -n -ŋ
1. S.D.	4n 4n 4ŋ	4n 4n 4ŋ	4m 4n 4ŋ	3m 1m 3n 1ŋ 4ŋ	4m 4n 4ŋ
2. J.W.	1m 3n 3n 2n 1ŋ 2ŋ	2m 2ŋ 4ŋ 4ŋ	4m 4n 4ŋ	4m 1m 2n 1ŋ 4ŋ	4m 3n 1ŋ 4ŋ
3. P.L.	4n 3n 1n 1ŋ 3ŋ	2m 1m 1n 2n 1ŋ 1ŋ 4ŋ	4m 4n 4ŋ	3m 1n 4n 4n	4m 4n 4ŋ
4. S.G.	1m 4n 3n 1n 3ŋ	2m 2n 2n 1n 2ŋ 3ŋ	4m 4n 4ŋ	4m 4n 3n 1ŋ	4m 1m 4n 1n 2ŋ
5. R.J.	2m 1n 1n 1n 1ŋ 3ŋ 3ŋ	2m 2n 2n 2ŋ 4ŋ	4m 4n 4ŋ	2m 2n 2ŋ 2ŋ 4ŋ	4m 1m 3n 1n 1ŋ 2ŋ
6. I.M.	1m 3n 3n 1ŋ 4ŋ	2m 4n 2n 4ŋ	4m 4n 4ŋ	3m 1m 3n 4ŋ	4m 1m 4n 3ŋ
7. V.P.	1m 3n 4n 1n 3ŋ	1m 3n 4n 4ŋ	4m 4n 4ŋ	4m 1m 3n 4ŋ	4m 4n 4ŋ
8. A.W.	2m 2n 1n 3ŋ 4ŋ	1m 2n 2n 1ŋ 2ŋ 4ŋ	4m 4n 4ŋ	4m 2m 2n 4ŋ	4m 1m 1n 3ŋ 3ŋ

SPEECH SIGNAL BY L.G.
RATIO OF NOISE LEVEL TO SPEECH SIGNAL: -18

STIMULI:	i	e	a	u	o
SUBJECTS	-m -n -ŋ	-m -n -ŋ	-m -n -ŋ	-m -n -ŋ	-m -n -ŋ
1. S.D.	3m 1n 4n 4ŋ	4m 4n 4ŋ	4m 4n 4ŋ	4m 4n 4ŋ	4m 4n 4ŋ
2. J.W.	2m 2n 4n 4ŋ	4m 2n 2ŋ 4ŋ	4m 4n 2n 2ŋ	4m 4n 4ŋ	4m 4n 4ŋ
3. P.L.	1m 3n 3n 1n 1ŋ 3ŋ	4m 3n 1ŋ 4ŋ	4m 4n 4ŋ	4m 4n 3n 1ŋ	4m 4n 4ŋ
4. S.G.	2m 2n 4n 1n 3ŋ	4m 4n 4ŋ	4m 4n 4ŋ	4m 1m 4n 3ŋ	4m 4n 4ŋ
5. R.J.	3m 1n 4n 2n 2ŋ	4m 4n 4ŋ	3m 1n 4n 4ŋ	4m 4n 4ŋ	4m 4n 4ŋ
6. I.M.	4m 4n 1n 2ŋ	4m 4n 3n 1ŋ 4ŋ	3m 4n 1ŋ 4ŋ	4m 1m 3n 4ŋ	4m 4n 1n 3ŋ
7. V.P.	2m 2n 4n 1n 3ŋ	4m 4n 4ŋ	4m 4n 1n 3ŋ	4m 1m 3n 4ŋ	4m 4n 4ŋ
8. A.W.	4m 4n 4ŋ	4m 4n 4ŋ	4m 4n 4ŋ	4m 4n 1n 3ŋ	4m 4n 4ŋ

Table 1b. Responses by individual subject to stimuli by Speaker L.G.

	-6 db			-12 db			-18 db			Total Cases
	-m	-n	-ŋ	-m	-n	-ŋ	-m	-n	-ŋ	
im →	2	30	0	9	22	1	8	24	0	96
in →	0	31	1	0	31	1	0	31	1	96
iŋ →	0	28	4	0	12	20	0	8	24	96
em →	2	30	0	9	23	0	7	25	0	96
en →	0	28	4	1	30	1	0	30	2	96
eŋ →	0	7	25	0	5	27	0	2	30	96
am →	25	6	1	27	5	0	29	3	0	96
an →	0	32	0	0	32	0	0	32	0	96
aŋ →	0	0	32	0	0	32	0	0	32	96
um →	18	10	4	31	1	0	28	4	0	96
un →	9	23	0	5	25	2	0	32	0	96
uŋ →	5	9	18	2	9	21	2	2	28	96
om →	18	3	11	25	3	4	27	5	0	96
on →	3	23	6	0	32	0	1	31	0	96
oŋ →	2	12	18	1	13	18	1	12	19	96

Table IIa. Responses by 8 subjects to stimuli by Speaker G.P.

	-6 db			-12 db			-18 db			Total Cases
	-m	-n	-ŋ	-m	-n	-ŋ	-m	-n	-ŋ	
im →	6	24	2	7	24	1	21	11	0	96
in →	0	17	15	1	22	9	0	31	1	96
iŋ →	1	16	15	0	6	26	1	6	25	96
em →	1	22	9	10	18	4	28	4	0	96
en →	0	19	13	3	18	11	0	28	4	96
eŋ →	0	3	29	0	1	31	0	1	31	96
am →	32	0	0	32	0	0	32	0	0	96
an →	0	31	1	0	32	0	0	32	0	96
aŋ →	1	3	28	0	0	32	0	3	29	96
um →	19	6	7	28	1	3	30	1	1	96
un →	19	6	7	6	23	3	1	31	0	96
uŋ →	0	3	29	0	7	25	1	3	27	96
om →	31	0	1	32	0	0	32	0	0	96
on →	1	23	8	0	27	5	0	32	0	96
oŋ →	5	2	25	4	2	26	1	0	31	96

Table IIb. Responses by 8 subjects to stimuli by Speaker L.G.

	-6 db			-12 db			-18 db			Total Cases
	-m	-n	-ŋ	-m	-n	-ŋ	-m	-n	-n	
im →	8	54	2	16	46	2	29	35	0	192
in →	0	48	16	1	53	10	0	62	2	192
iŋ →	1	44	19	0	18	46	1	14	49	192
em →	3	52	9	19	41	4	35	29	0	192
en →	0	47	17	4	48	12	0	58	6	192
eŋ →	0	10	54	0	6	58	0	3	61	192
am →	57	6	1	59	5	0	61	3	0	192
an →	0	63	1	0	64	0	0	64	0	192
aŋ →	1	3	60	0	0	64	0	3	61	192
um →	37	16	11	59	2	3	58	5	1	192
un →	28	29	7	11	48	5	1	63	0	192
uŋ →	5	12	47	2	16	46	3	6	55	192
om →	49	3	12	57	3	4	59	5	0	192
on →	4	46	14	0	59	5	1	63	0	192
oŋ →	7	14	43	5	15	44	2	12	50	192

Table III. Responses by 8 subjects to stimuli by Speaker G.P. and Speaker L.G.

Based on Table II, Figure 2 was made. It shows the misidentification of syllable final nasal consonants m, n, ŋ, in each VC syllable type in percent. The percentages are presented with respect to the three noise conditions. It should be noted that the scale of the figures shown in Figure 2 differs according to the number of misidentified stimuli in each case.

These results suggest three principal findings: (1) [-m] tends to be identified as [-n] after the front vowels [i] and [e] as shown in Figure 2c; (2) [-ŋ] tends to be identified as [-n] after vowel [i]; and (3) [-m], [-n] and [-ŋ] tend to be correctly identified after vowel [a] even in the noisiest condition.

In addition to these three principal findings, there are some tendencies which are less general in the sense that they are observable only in the speech of one of the two speakers. For instance, as shown in Figure 2d, [-ŋ] has a tendency to be identified as [-n] after vowels [u] and [o]. However, this tendency can be observed only in the speech of Speaker G.P., and not in the speech of Speaker L.G. As shown in Tables IIa and IIb, the number of responses for [-ŋ] being identified as [-n] after vowels [u] and [o] for the speech of Speaker G.P. are 9, 9, 2 (for the three noise conditions respectively) and 12, 13, 12, but 3, 7, 3 and 2, 2, 0 for the speech of Speaker L.G. Similarly, [-n] has a tendency to be identified as [-ŋ] after vowels [i] and [e], as shown in Figure 2f. The number of responses for [-n] being identified as [-ŋ] after [i] and [e] are greater for the speech of Speaker L.G. (15, 9, 1 and 13, 11, 4) than for the speech of Speaker G.P. (1, 1, 1 and 4, 1, 2) as shown in Tables IIa and IIb. In fact, in Speaker L.G.'s speech, [in]/[iŋ] are essentially confused with each other.

A third type of finding concerns tendencies which are weak but are represented in the speech of both speakers. Thus [-n] has a tendency to be identified as [-m] after vowel [u], as shown in Figure 2a. It is more obvious in the speech of Speaker L.G. (19, 6, 1) than in the speech of Speaker G.P. (9, 5, 0), as shown in Tables IIa and IIb.

Figure 3a shows the percentage of correct identifications of the nasal consonants regardless of both their vowel environments and the noise conditions. [-n] has the highest percentage of being correctly identified, to be followed by [-ŋ], and [-m] has the lowest percentage.

Figure 3b shows the percentage of the correct identifications of all the nasal consonants as a function of the vowel environments. The most favorable environment for correct identification is [a-] and the remaining vowel environments rank in order of decreasing identifiability as follows: [o-], [u-], [e-] and [i-].

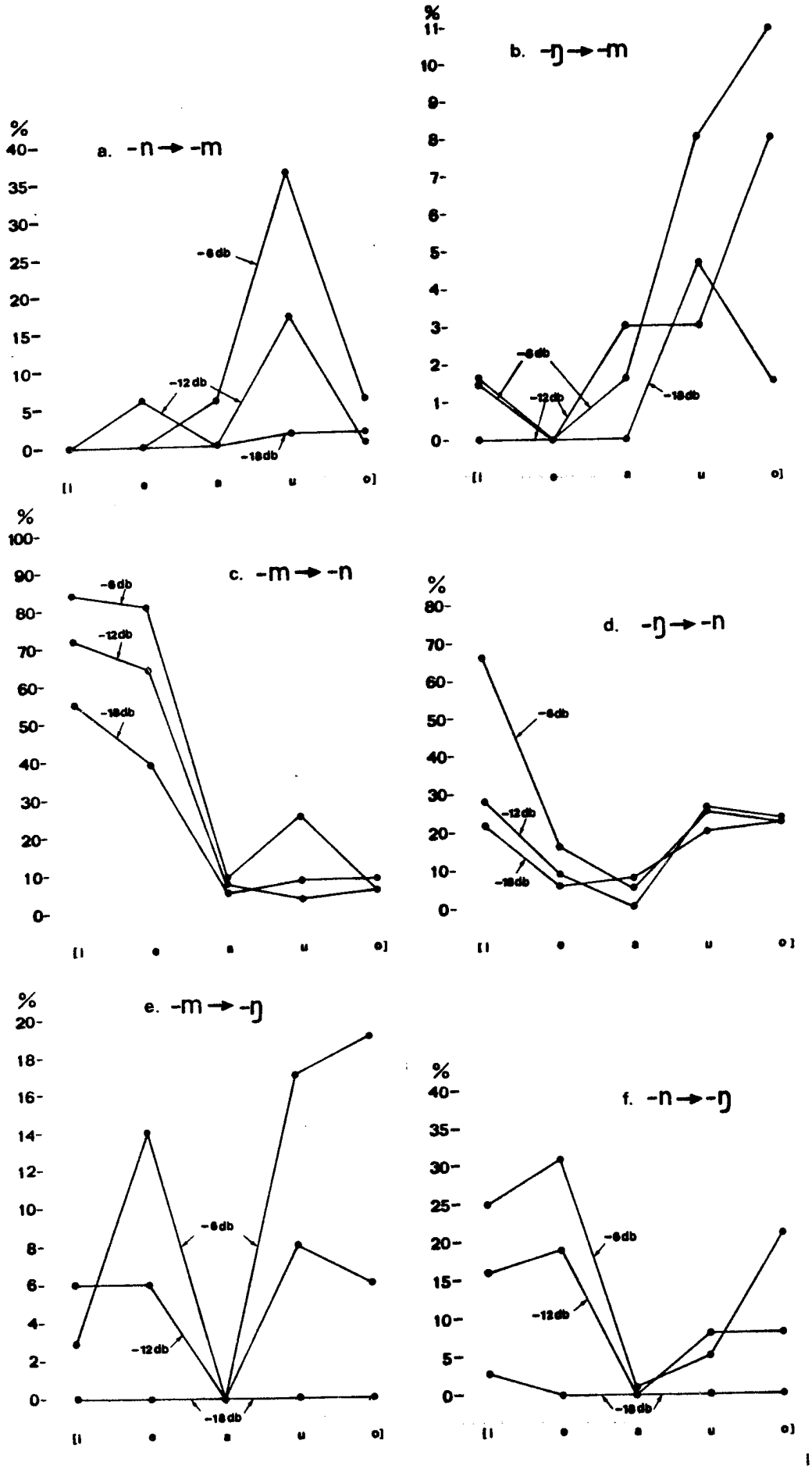
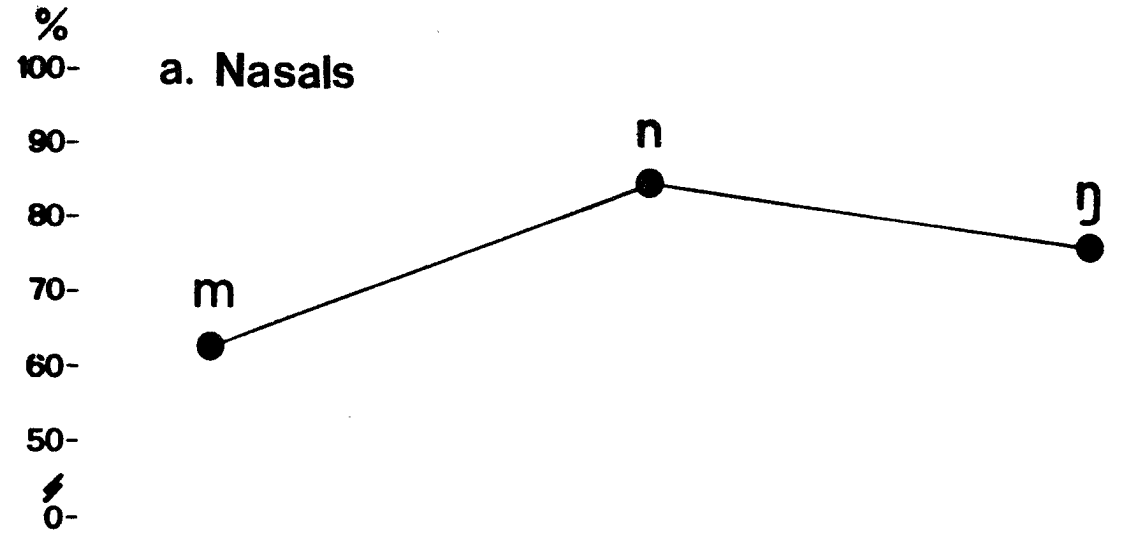
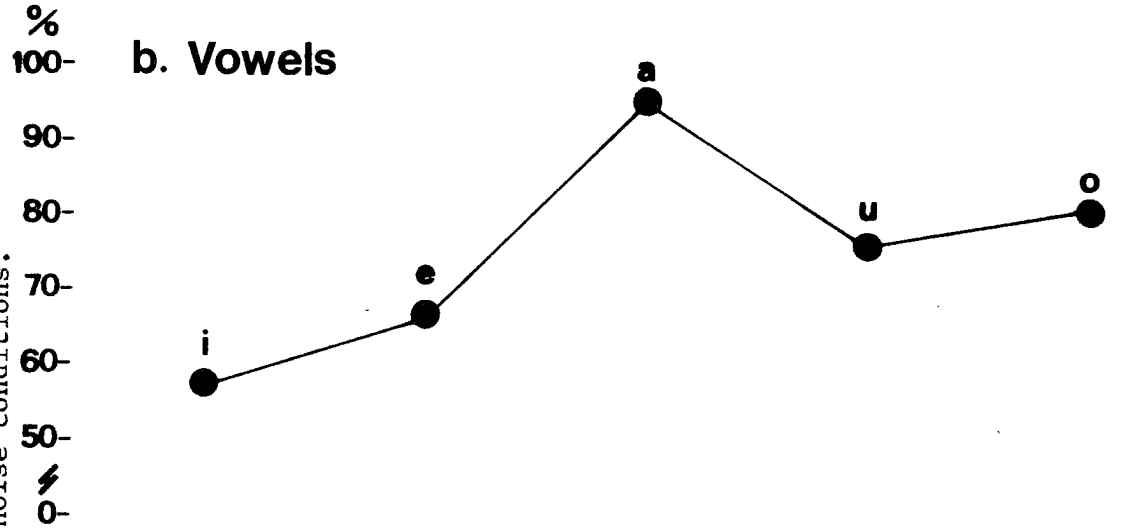


Figure 2. The percentages of misidentifications of syllable final nasal consonants [-m], [-n] and [-ŋ] in each VC syllable type with respect to the three noise conditions.





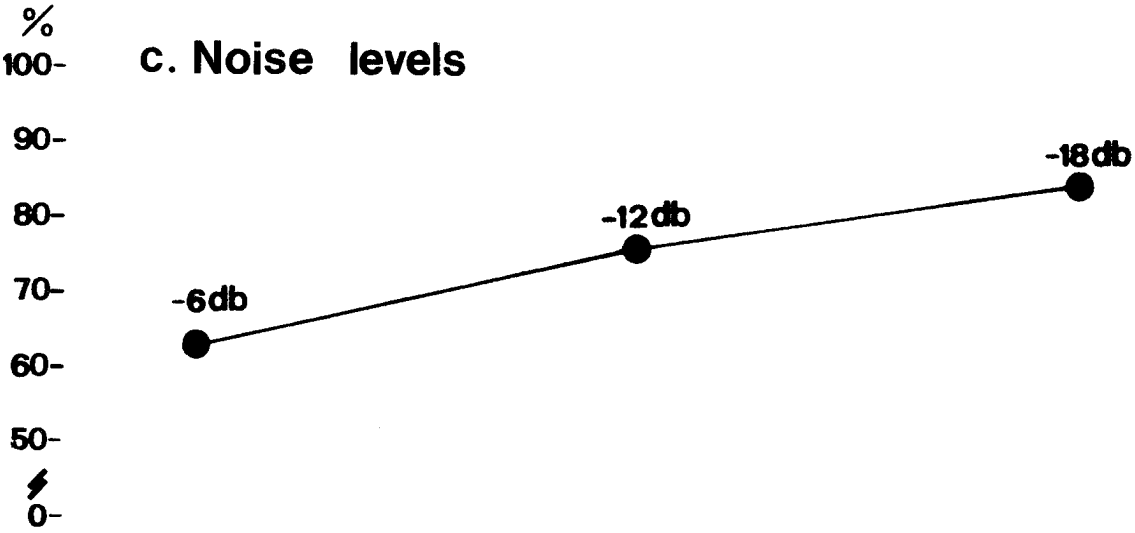


Figure 3. The percentages of correct identifications of syllable final nasal consonants [-m], [-n] and [-ŋ] (a) regardless of vowel environments and noise conditions; (b) as a function of vowel environments; (c) as a function of noise conditions.

Figure 3c shows the percentage of correct identifications of the nasal consonants as a function of the noise conditions. As expected, the percentage increases as noise is reduced.

In order to determine the degree of significance of the effect of the vowel, the speaker or the noise condition on the misidentificaion of the nasal consonants, a one-way analysis of variance was done. The 3 independent variables include the vowels, the noise conditions, and the speakers, and the main effects are vowel, noise and speaker. There are six types of misidentification of the consonants possible, namely,

Vn → Vm	Vm → Vn	Vm → Vŋ
Vŋ → Vm	Vŋ → Vn	Vn → Vŋ

(V = vowels)

The result of the analysis of variance is shown in Table IV. Vowel as one of the main effects is highly significant ($p < 0.001$) for all the cases of confusion, which indicates that vowels in fact affect the perception of the syllable final nasal consonants. Except for the case of Vŋ → Vn, noise level is also highly significant ($p < 0.001$). The speaker effect is highly significant ($p < 0.001$) for the cases of Vm → Vn, Vn → Vŋ and Vŋ → Vm, and not significant for Vn → Vm, Vŋ → Vn, and Vm → Vŋ. This may cautiously be taken as an indication that the pattern of misidentifications in the latter instances are similar with both speakers. Where there is a significant speaker effect it does not necessarily mean that the given type misidentification is not found with both speakers but may indicate that it is more prevalent with one of the speakers.

4. DISCUSSION I.

As displayed in the spectrograms of syllables [im] and [in] (Figures 4 and 5), the only noticeable difference, as far as formant transition is concerned between these two syllables, is in F2. For both Speaker G.P. (Figure 4) and Speaker L.G. (Figure 5), both F2 and F3 for [im] and [in] shift downward at the transition, however, the degree of shift in F2 transition is greater for [im] than for [in]. We can reasonably assume that this greater shift in F2 transition in [im] plays an important role in differentiating [im] from [in], since this is the only noticeable difference between the syllables. Possibly the presence of noise selectively attenuates the perceptability of the transition from vowel to consonant and the stimulus is perceived as equivalent to one with the smallest transition, i.e., as [in]. Notice that, as shown in Table IIa, in the less noisy conditions (-12 db and -18 db), the number of correct identifications of [im] for the speech of Speaker G.P. does not improve much over the noisiest condition. However, for the speech of Speaker L.G., as shown in Table IIb, there is an appreciable

ANALYSIS OF VARIANCE
(One way)

INDEPENDENT VARIABLES:

1. VOWELS (5)
2. MASKING NOISE (3)
3. SPEAKERS (2)

RESULT:

MISIDENTIFICATION	MAIN EFFECTS	F-RATIO	df_1	df_2 (=N-df_1+1)	SIGNIFICANCE
Vm→Vn	Vowel	176.56	4	300	$p < 0.001$
	Noise	24.12	2	302	$p < 0.001$
	Speaker	72.13	1	303	$p < 0.001$
Vm→Vŋ	Vowel	5.22	4	44	$p < 0.001$
	Noise	20.17	2	46	$p < 0.001$
	Speaker	2.27	1	47	$p < 0.132$
Vn→Vm	Vowel	27.44	4	45	$p < 0.001$
	Noise	16.92	2	47	$p < 0.001$
	Speaker	0.11	1	48	$p < 0.744$
Vn→Vŋ	Vowel	15.61	4	90	$p < 0.001$
	Noise	22.37	2	92	$p < 0.001$
	Speaker	35.23	1	93	$p < 0.001$
Vŋ→Vm	Vowel	26.69	4	22	$p < 0.001$
	Noise	23.35	2	24	$p < 0.001$
	Speaker	8.10	1	25	$p < 0.001$
Vŋ→Vn	Vowel	6.03	4	171	$p < 0.001$
	Noise	2.35	2	173	$p < 0.096$
	Speaker	0.17	1	174	$p < 0.676$

(V = vowels ; N = number of cases of misidentification)

Table IV. Result of the analysis of variance.

increase in the correct identifications of [im]. This may be due to the fact that the difference in F2 tansition between [im] and [in] for Speaker G.P. is small, as shown in Figure 4, whereas there is a greater difference in the F2 transitions of [im] and [in] in the speech of Speaker L.G., as shown in Figure 5. It is this difference between the speakers that suggests that there may be a contribution to these errors from masking of F2 transitions. As there is less difference between [im] and [in] for Speaker G.P. in this respect, the reduction of noise level provides less improvement in identification than it does for Speaker L.G.

Nasal murmur quality may also play a role in the [im] → [in] misidentifications. House (1957) synthesized nasal consonants by coupling a nasal tract analog to a vocal tract analog. For a bilabial nasal consonant, the cross-sectional area of the vocal tract was zero at the end of the vocal tract. The configurations for the vocal tract analog were appropriate for the production of vowels i, a and u. Added output of the nasal tract analog and the vocal tract analog were produced for bilabial nasal consonants with the vowel configurations for i, u and a. These bilabial nasal consonants were used as stimuli for a listening test. A crucial finding was that the bilabial nasal m with the configuration for i produced an output that was heard as n more than half of the time. In natural speech, the production of a syllable consisting of a vowel followed by a nasal will most probably involve the process of perseverative coarticulation between the two sounds. Coarticulation of this type may be related primarily to mechanicoinertial factors which cause articulator response to lag behind the arrival of neural commands and to persist after such commands cease (Daniloff and Moll, 1968, p. 708; Lindblom, 1963; Stevens and House, 1963). In our case, the vocal tract configuration for the production of vowel [i] may still be maintained during the production of the succeeding [m]. As a result of perseverative coarticulation, the murmur quality of [m] may be approximated to the murmur quality of [n], as the vocal tract configuration for the production of [m] is now constricted in the region appropriate for the production of [i]. The effect is presumed to be that the distance from the glottis to the constriction in the oral cavity is more like that typical for [n] than for [m]. The modified murmur quality of [m] in [im] may thus be another factor causing the [im] → [in] misidentifications. It is possible that an obscured F2 transition and the modified nasal murmur quality both contribute to the [im] → [in] misidentifications, although it is not possible to say which is more responsible than the other. For example, Nord (1976) found it impossible to quantify relations between nasal resonance and formant transition with respect to the identification of [m] and [n].

The tendency for [em] to be identified as [en] may be explained in a similar fashion, since the difference in formant transition pattern between [em] and [en] is similar to that between [im] and [in], as shown

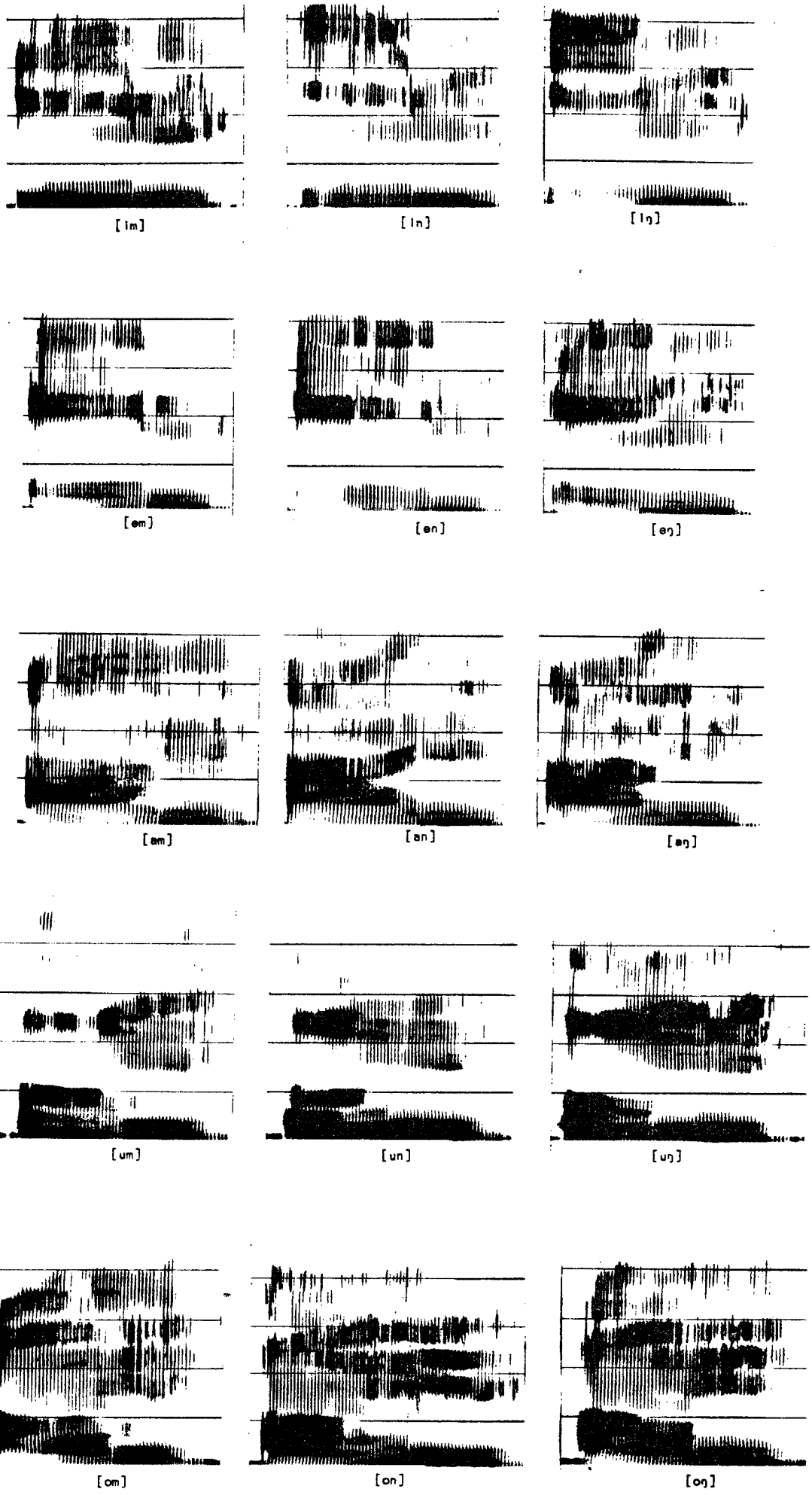


Figure 4. Spectrograms of [im], [in], [iŋ], [em], [en], [eŋ], [am], [an], [aŋ], [um], [un], [uŋ], [om], [on], [oŋ] produced by Speaker C.P.

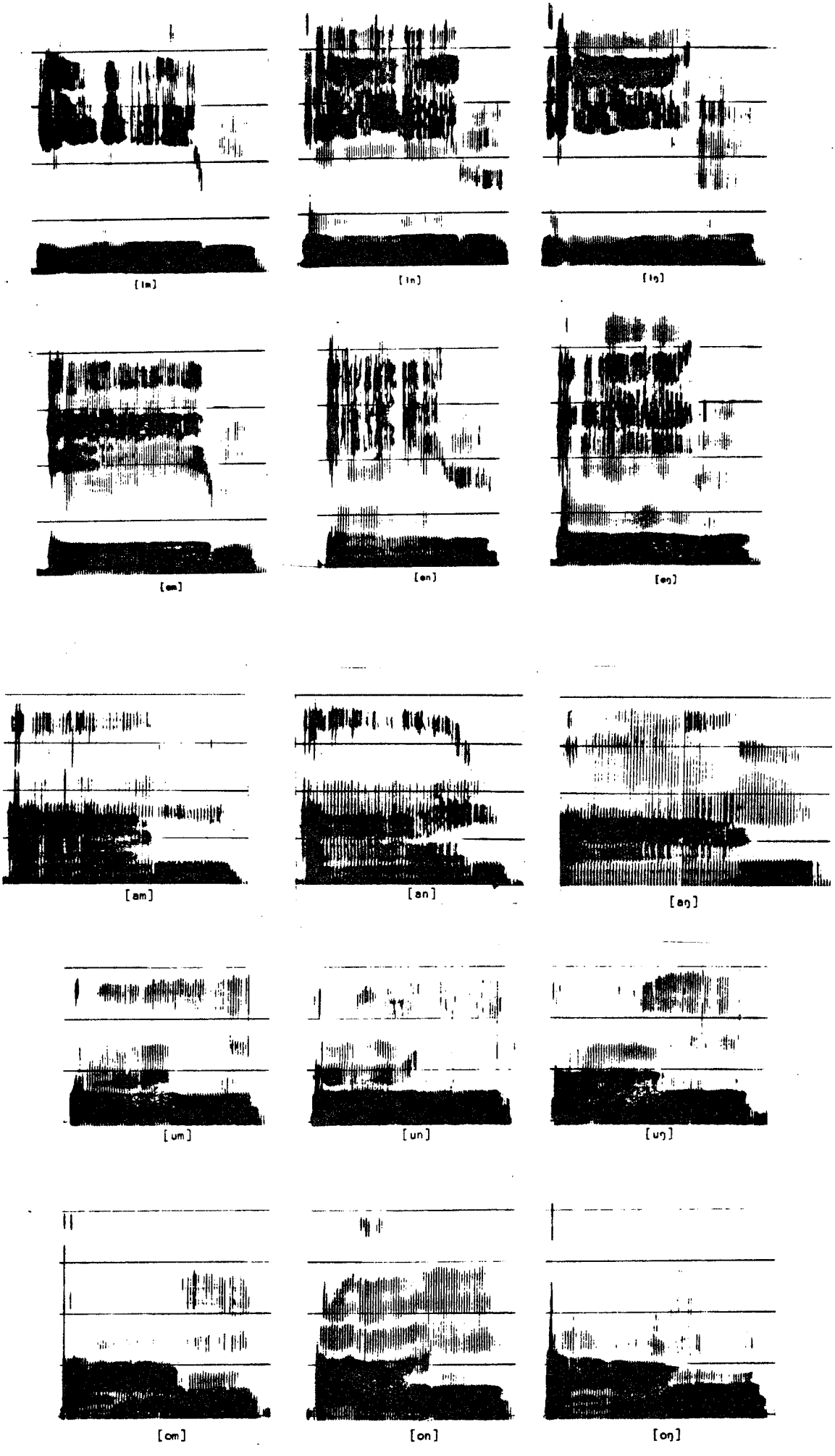


Figure 5. Spectrograms of [im], [in], [iŋ], [em], [en], [eŋ], [am], [an], [aŋ], [um], [un], [uŋ], [om], [on], [oŋ] produced by Speaker L.G.

in the spectrograms of [em] and [en] in Figure 4 for the speech of Speaker G.P. and Figure 5 for the speech of Speaker L.G. Furthermore, the difference in F1 and F2 values between [e] in [em] and [i] in [im] is not substantial. For the speech of Speaker G.P. (Figure 4), the values of F1 and F2 are roughly estimated as 300 Hz and 2350 Hz for vowel [i] and 400 Hz and 2200 Hz for vowel [e]; and for the speech of Speaker L.G. (Figure 5), the F1 and F2 values are roughly estimated as 250 Hz and 2500 Hz for vowel [i] and 350 Hz and 2250 Hz for vowel [e]. Similar to the [im]→[in] misidentifications, again there is a considerable decrease of misidentifications of [em] as [en] for the speech of Speaker L.G. as the noise level decreases (Table IIb, but not for the speech of Speaker G.P. (Table IIa). Again, this may be due to the fact that the difference in F2 transition between [em] and [en] is much greater for the speech of Speaker L.G., than for the speech of Speaker G.P. (Cf. Figures 4 and 5).

What may be the possible cause of the [iŋ]→[in] misidentifications? Compare the formant transition patterns in [iŋ] and [in], as shown in the spectrograms in Figure 4 and Figure 5. Unlike those in [in], both F2 and F3 in [iŋ] do not shift downward but maintain level at the transition for both speakers, and in the speech of Speaker L.G. (Figure 5), F4 is shifting upward at the transition. Since the masking noise cannot cause a downward shift in F2 and F3 at the transition, a loss in the perceptability of the transitions cannot explain why [ŋ] is heard as [n] after [i]. A possible explanation for the [iŋ]→[in] misidentifications would seem to be the nasal murmur quality of [ŋ] in [iŋ]. It is likely that the oral closure for the production of [ŋ] may be shifted forward toward the palatal region of the vocal tract as a result of the perseverative coarticulation effect of the preceding vowel [i]. The fronting of the closure may cause the murmur quality of [ŋ] to change, as the distance between the closure and the glottis is increased compared to a truly velar [ŋ]. A fronted [ŋ] would sound to a certain degree similar to a palatal nasal consonant [ɲ]. Since [ɲ] was not a given response category for the subjects, [n] would seem to a reasonable, or close, substitute.

The subjects' responses in this task may also have been influenced in certain respects by the relative frequency of the different nasals in English and the relative frequency of different vowel-nasal sequences. Responses were compared with estimates of frequencies of 2-phoneme sequences prepared by Denes (1963) and Carterette and Jones (1974). Denes investigated spoken Standard Southern British English, using a body of conversational material and narrative taken from Phonetics Readers, including 72,210 phonemes, forming 29,916 syllables, and 23,052 words. The frequencies of occurrence of digrams which concern us here are listed as follows:

		2nd phoneme		
		m	n	ŋ
1st phoneme	i	38 (26%)	113 (73%)	1 (1%)

Carterette and Jones calculated the frequencies of occurrence of digrams in a text spoken by adult speakers of Californian English. The frequencies which concern us here are:

		2nd phoneme		
		m	n	ŋ
1st phoneme	i	45 (24%)	123 (75%)	1 (1%)

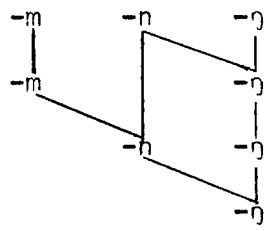
In Both cases, we can see that the frequency of occurrence is greater for the digram, in, than for the digrams, im and iŋ. In our experiment, subjects responded [in] in 65% of occurrences [im], [in] and [iŋ], whereas [im] was only responded 10% of the time. A response bias in favor of [in] would partly explain these results. However, [iŋ] was the response in 25% of the cases despite the rarity of this sequence in English. With this vowel at least, the results do not seem to be explicable as primarily due to response bias.

One of our principal findings is that the nasal consonants, [m], [n], [ŋ], following vowel [a] tend to be correctly identified even in the noisiest condition. This may be explained in terms of formant transition, intrinsic vowel intensity and nasal murmur quality. Compare the difference in formant transition among the syllables [am, an, aŋ] (See Figure 4 for the speech of Speaker G.P., and Figure 5 for the speech of Speaker L.G.) with the same difference among other syllable types: [im, in, iŋ], [em, en, eŋ], [um, un, uŋ] and [om, on, oŋ] (See also Figure 4 and Figure 5 for Speaker G.P. and Speaker L.G. respectively). We can see that the differences in the formant transitions after [a] are far more conspicuous than differences after any of the other vowels for the speech of both speakers. The more conspicuous the difference in formant transition is, the more salient the perceptual distinction among the nasal consonants is likely to be, and it follows that these nasal consonants are more likely to be correctly identified. Furthermore, since the vowel [a] has the greatest intrinsic vowel intensity (Lehiste, 1970) compared to the other vowels, the formant transition in [am], [an] or [aŋ] should be more resistent to masking noise and is thus less likely to be obscured. Furthermore, [a] is least likely to cause any perseverative coarticulation effect on a following nasal consonant, whether it

is [m], [n] or [ŋ], as [a] is a low vowel which is produced with the body of the tongue depressed and without forming a constriction above the pharyngeal area in the vocal tract similar to that which occurs in any of the nasal consonants. Thus, nasal murmur of [m], [n] or [ŋ] is unaffected by the environment [a-]. Given these factors, it is not surprising that the nasal consonants are most identifiable when they occur after vowel [a].

5. DISCUSSION II.

Our result may provide some understanding of diachronic change in syllable final nasal consonants, especially in a monosyllabic language, such as Chinese. Since Ancient Chinese, a language around 600 A.D. (Karlgren, 1954), the syllable final nasal consonants, -m, -n, -ŋ, have undergone a process of merging. Chen (1972a, 1973) claims that the nasal consonants first merged to two than to a single nasal final, as shown below:



The direction of merging, according to Chen, is invariably from the front to the back along the dimension of the place of articulation: -m > -n, -n > -ŋ (although -m > -ŋ did not take place in one step). The questions to be raised are, why did the merger occur in this direction, but not the reverse, such as -n > -m, -ŋ > -n or -ŋ > -m? Why did not a direct merger such as -m > -ŋ occur?

Based on Table III, we obtain the total number of the cases of misidentification as shown in the following:

Misidentification cases	Number of cases	
Vm → Vn	304	
Vŋ → Vn	175	
Vn → Vŋ	94	(V = vowels)
Vn → Vm	49	
Vm → Vŋ	48	
Vŋ → Vm	26	

These data show that it is not a matter of front-to-back or back-to-

front misidentifications being more or less likely. The most common misidentification, [-m] → [-n] is front-to-back, but the next most likely, [-ŋ] → [-n], is back-to-front. Those cases where only a small number of misidentifications occur in our perceptual data would seem to predict that changes of such as -n > -m, -ŋ > -m and -m > -ŋ would not occur in syllable final nasals. But the comparatively high misidentification rates would predict that changes such as -m > -n and -n > -ŋ have occurred. Yet, according to Chen's proposal, the change -ŋ > -n did not occur in the merging process. However, there are cases of the change -ŋ > -n in many major Chinese dialects. For example, in Hakka and Szechwan (Karlgren, 1915-1926), the following cases actually occurred:

	600 A.D.		1926	
Hakka dialect:	$k^hj\iota\varepsilon ŋ$	>	k^hin	'light'
	$kj\iota\varepsilon ŋ$	>	kiaŋ	'neck'
Szechwan dialect:	$ji^w\text{ɒ}ŋ$	>	yin	'forever'
	$xji^w\text{ɒ}ŋ$	>	ɕioŋ	'older brother'

Interestingly enough, both in Hakka and Szechwan, the syllable final velar nasal does not change to -n, unless the preceding vowel has changed to a high front vowel, i. Similar cases occurred in numerous dialects in the provinces of Yunnan (Yang, 1969), Hubei (Chao, et al, 1948), Hunan (Yang, 1974) and Hebei (Peking Normal College, 1961). This correlates with the fact that the [iŋ] → [in] misidentification is perceptually more probable, which is one of our principal findings. Thus, the result of our study has provided a perceptual basis for explaining why in the process of change in syllable final nasal consonants in the dialects of Chinese such cases as *-n > -m, *-ŋ > -m, *-m > -ŋ did not occur and why such a case as -ŋ > -n should and did in fact occur.

6. CONCLUSION AND SUMMARY OF RESULTS.

We conclude that vowel quality in fact affects the perception of syllable final nasal consonants, and our results have provided a perceptual basis for explaining the direction of the diachronic change in the syllable final nasal consonants in the dialects of Chinese. Our principal findings and other results are summarized as follows:

(1) [-m] tends to be identified as [-n] after the front vowels [i] and [e].

(2) [-ŋ] tends to be identified as [-n] after the front vowel [i].

(3) [-m], [-n] and [-ŋ] tend to be identified correctly after the vowel [a] even in the noisiest condition.

(4) As a function of both their vowel environments and the noise conditions, [-n] has the highest percentage of being correctly identified, followed by [-ŋ]; [-m] has the lowest percentage.

(5) There is a weak tendency for [-n] to be identified as [-m] after the vowel [u].

ACKNOWLEDGMENT

The author wishes to thank Steve Anderson, Vanna Condax, Susie Curtiss, Sandy Disner, Vivian Flores, Manuel Godinez, Jr., Louis Goldstein, Steve Greenberg, Richard Janda, Peter Ladefoged, Wendy Linker, Ian Maddieson, George Papcun, Ann Wingate, Jim Wright and Jenie Yamada for their participation in the experiment. The author is especially grateful to Ian Maddieson and Peter Ladefoged for reading an earlier version of this paper and suggesting many valuable ideas, to Vincent Van Heuven for his assistance in statistical analysis, and to Ronold Carlson, Steve Geenberg and Willie Martin for their technical assistance. [Research supported by NSF]

REFERENCES

Carterette, E.C. & M.H. Jones (1974). Informal Speech. Los Angeles, Univ. of Calif. Press.

Chao, Y.R. & others (1948). Report on a Survey of the Dialects in Hubei Province. Taipei, Institute of History and Philology.

Chen, Matthew (1972). Nasals and Nasalization in Chinese: Explorations in Phonological Universals. Ph.D. Diss. (UC, Berkeley).

Chen, Matthew (1973a). Cross-dialectal comparison: a case study and some theoretical considerations. J. of Chinese Linguistics, 1, 38.

Daniloff, R. & K. Moll (1968). Coarticulation of lip-rounding. J. of Speech Hearing Research, 11, 707.

Denes, P.B. (1963). On the statistics of spoken English. JASA, 35, 892.

Karlgren, B. (1915-1926). Etudes sur la phonologie chinoise. Stockholm, Norstedt & Söner.

Karlgren, B. (1954). Compendium of phonetics in Ancient and Archaic Chinese. Stockholm, the Museum of Far Eastern Antiquities, Bulletin, 26, 211.

Lehiste, I. (1970). Suprasegmentals. Cambridge, MIT Press.

Liberman, A.M. and others (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychological Monographs, 68, #8, 1.

Lindblom, B. (1963). Spectrographic study of vowel reduction. JASA, 35, 1773.

Malécot, A. (1956). Acoustic cues for nasal consonants: an experimental study involving a tape-splicing technique. Language, 32, 274.

Nakata, K. (1959). Synthesis and perception of nasal consonants. JASA, 31, 661.

Nord, L. (1976). Perceptual experiments with nasals. STL-QPSR, 2-3, 5.

Peking Normal College (1961). A Description of the Dialects in Hebei Province. Tienjin, Hebei Renmin Chubanshe.

Stevens, K.N. & A. House (1963). Perturbation of vowel articulations by consonantal context: an acoustical study. J. of Speech Hearing Research, 6, 111.

Yang, S.F. (1969). A Report on the Survey of the Dialects in Yunnan Province. Taipei, Institute of History and Philology (Special Publication, 56).

Yang, S.F. (1974). A Report on the Survey of the Dialects in Hunan Province. Taipei, Institute of History and Philology (Special Publication, 66).

On the perception of contour tones

Steven Greenberg and Eric Zee

[Revised version of a paper presented at the 94th meeting of the Acoustical Society of America, Miami Beach, Florida, 12-16 December, 1977]

I. INTRODUCTION

F_o (and its perceptual correlate pitch)[1] plays an important and multifaceted role in the perception of speech. Among its many functions in English and numerous other languages it (1) serves to distinguish sentence types (Lieberman, 1965), (2) contributes to the delineation of syntactic boundaries (Fodor, Bever, and Garret, 1974), (3) figures in stress assignment (Fry, 1955), and (4) serves as a voicing cue (Haggard, Ambler and Callow, 1970). However, the function of F_o most relevant in the present experimental context is one rarely found in the Indo-European language family but quite commonly encountered among the Sino-Tibetan languages of East and Southeast Asia.

In the languages of this area the F_o pattern - or tone contour - serves to invest lexical items composed of identical phonological segments[2] with entirely distinctive meanings. Thus in Mandarin Chinese there are monosyllabic morphemes of the form ma with four meanings depending on the tone carried:

		Tone Height Pattern[3]	Tone Type	Meaning
(a)	ma^1	55	high level	"mother"
(b)	ma^2	35	high rising	"hemp"
(c)	ma^3	214	low falling rising	"horse"
(d)	ma^4	51	high falling	"scold" (Ladefoged, 1975:228)

[1] The fundamental need not be present for the perception of a basic pitch to occur (Klatt, 1973)

[2] The phonological level is for the purposes of the present discussion excludes tonological units.

[3] Within the present context, 1 represents the lowest pitch and 5 the highest.

The range of lexical tone patterns, though theoretically infinite is, in fact, quite small. Few lexical tone languages have more than (a) four contour tones, (b) two rising tones, (c) two falling tones, or (d) two convex tones (Wang, 1972). Moreover, no tone language has more than five distinctive pitch levels (Wang, 1967).

A variety of operational constraints affecting the perceptual and articulatory mechanisms underlying speech can influence the structure of tonemic systems. The operation of the larynx is governed by a combination of mechanical and physiological factors that places limits on the acceleration and deceleration of vocal fold vibration (Sundberg, 1973; Ewan, 1976). The auditory system functions within a comparable set of limits concerning its capacity to spectrally resolve the components of different tones. Consequently, the range of vocal pitch remains fairly constant across languages regardless of the number of tones a language has (Wang, 1967)[4]

The present study sought to determine whether certain general properties of contour tone perception can account for the asymmetrical pattern of tonemic distribution present in many dialects of Chinese. In dialects such as Shanghai and Taiwanese, the tone system is divided into two categories - long tones (170 - 350 msec) and short tones (70 - 120 msec) - on the basis of duration. The tones are also classifiable according to whether they are level or contour. Zee and Greenberg (1977) observed that short tones in these dialects are tonemically classified as level tones despite their acoustic resemblance to the long contour tones (Figure 1). The disparity between the physical features and linguistic (tonemic) classification of the short tones raises certain questions concerning the contribution of duration to the perception of tonal contouricity. Are the short (physically contour) tones perceived as being level or contour? If the short tones are perceived as level, can the percept be altered by lengthening the initial segment duration?

Experiment I is concerned with some general features of auditory mechanisms engaged in the processing of contour tones, namely the relationship of the rate and range of F_o change to the degree of perceived contouricity.

II. EXPERIMENT I: GENERAL FEATURES OF CONTOUR TONE PERCEPTION

Stimuli. Ten variants of the vowel [i] were digitally synthesized on a PDP-12 computer. Formants for the stimuli were set at the following center frequencies: F1=300, F2=2500, F3=3000, F4=3380 and F5=4165 Hz. A variable fundamental frequency contour was superimposed on this five-formant pattern as illustrated in Figure 2. The contour was divided into two parts, an initial portion consisting of an unmodulated fundamental frequency fixed at 100 Hz and a terminal ramp whose F_o ascended in quasi-linear fashion to an offset frequency ranging between 104 and 164 Hz.

---

[4] "Maddieson (p. c) suggests that the number of tones may affect the pitch range used if other things are equal".

The duration of the ramp varied between 4 and 16 pitch periods. Ramp duration varied slightly as a function of frequency range traversed. Durations for the 4 period ramp varied between 33.5 and 39 msec, the 8 period ramp ranged between 68.2 and 76.5 msec, and the 16 period ramp varied between 121.7 and 147.6 msec. The rate of frequency ascent varied between 1 and 8 Hz increments per pitch period.

The duration of the initial steady-state portion of the contour was roughly inversely proportional to ramp length. The 4 period ramps were preceded by a 16 cycle (160 msec) steady-state F_o; the 8 period ramps by 8 periods (80 msec) of a level F_o and the 16 period ramps by a 4 period (40 msec) initial segment.

The set of ten stimuli thus generated varied along two independent dimensions - ramp duration and rate of frequency change. The third dimension, size of the frequency range traversed by the ramp is simply a product of the other two. The composition of the stimulus set is summarized in Table I.

Method. The stimuli were presented over loudspeaker to a group of 12 native English-speaking individuals who donated their time in the interest of furthering human knowledge. The entire set of ten stimuli was played before proceeding to the experiment proper in order to allow the listeners to hear the complete range. In the experiment subjects were asked to rate each stimulus on a scale from 1 to 7 as to its degree of "contouricity" - defined within the experimental context as "the steepness of the rise in pitch" during the terminal portion of the vowel. For purposes of consistent reference "1" was defined as a completely level tone, "7" as a really steep contour, and "4" as a tone midway between "1" and "7".

There were 120 presentations of the ten stimuli, the first 20 of which were discarded in the analysis. The remaining 100 trials contained ten repetitions of each stimulus. The interval between successive stimuli was 3 sec.

Results. The degree to which each stimulus was perceived as a contour is presented in Table II. Multiple T-tests were performed to obtain a rough idea of the stimulus clustering pattern. Stimuli whose means were not significantly different at the .05 level were placed in the same group. On the basis of these statistical tests five groups were defined:

A	B	C	D	E
1	(3)[5]	6	7	10
2	4	8	9	
3	5			

[5] Stimulus 3 was classified as being in Group A for the purposes of the analysis although its rating did not significantly differ from that of stimuli 2 or 4.

TABLE I.

Stimulus Set

Rate of frequency change per pitch period

Ramp Durations (Pitch periods)				
4	ΔF = 4 Hz Stimulus 1	ΔF = 8 Hz Stimulus 2	ΔF = 16 Hz Stimulus 3	ΔF = 32 Hz Stimulus 4
8	ΔF = 8 Hz Stimulus 5	ΔF = 16 Hz Stimulus 6	ΔF = 32 Hz Stimulus 7	----------
16	ΔF = 16 Hz Stimulus 8	ΔF = 32 Hz Stimulus 9	ΔF = 64 Hz Stimulus 10	----------

TABLE II.

"Contouricity" Ratings for Experiment I

Stimulus

	1	2	3	4	5	6	7	8	9	10
$\bar{X}$	1.30	1.36	1.90	2.24	2.23	3.58	4.48	3.88	4.76	5.90
σ	.31	.36	.60	.86	.55	1.04	.86	.56	.60	.76

A consistent pattern of clustering emerges for groups C, D and E. Stimuli 6 and 8 of Group C and stimuli 7 and 9 of group D have identical frequency ranges (16 Hz and 32 Hz, respectively). Stimulus 10 with a frequency range twice as great as any other contour is classified in a group by itself. Group B, composed of stimuli 4 and 5 is united by no apparent common dimension. The frequency range subtended by stimulus 4 is four times as great as stimulus 5, while its rate of frequency modulation is eight times as great. The composition of group A provides a clue for the odd coupling of stimuli in Group B. The ramps of the stimuli in this group are all four periods long, suggesting that a ramp whose duration lies below a critical threshold is perceived as a level tone regardless of its degree of frequency modulation. It must be emphasized that the rates of change for even the least steep ramp are not inconsiderable. They are equivalent to 100 Hz/sec, 200 Hz/sec, and 400 Hz/sec for stimuli 1, 2, and 3, respectively. Stimulus 4, whose rate of change is equivalent to 800 Hz/sec is perceived as having a degree of contouricity equivalent to that of a ramp eight times less steep and covering only one fourth the frequency range (stimulus 5).

The data from Table II, as plotted in Figures 3(a) and 3(b) provides another view of the relationship between ramp duration and the degree of perceived "contouricity". It is evident from these figures that a related factor governs the growth of perceived contouricity for the 8 and 16 period ramps both as a function of rate of frequency change (Fig. 3a) and frequency ranged subtended (Fig. 3b). Interestingly the 4 period ramps have a markedly attenuated growth function. The similarity of the slopes for the 8 and 16 period stimuli tentatively suggest that a common growth function may apply to all contours exceeding a durational threshold lying tween 40 and 65 msec.

Discussion. That the sensation of dynamic pitch ("Contouricity") depends on a minimum signal duration of 40 - 65 msec may seem somewhat surprising in view of the fact that the precision of pitch discrimination for single frequencies begin to stabilize for durations somewhere above 12 msecs. (Dougherty and Garner, 1948). However, a number of studies suggest that the perception of dynamic pitch differs qualitatively from the processing of fixed frequencies. For example the masked threshold of ramped tones is 3 to 4 dB higher than for steady frequencies (Collins and Cullen, 1978). Moreover, the difference limen for discrimination of vowels with dynamic F_o patterns is an order of magnitude greater than for vowels with a steady F_o namely 2 Hz vs. .3 Hz (Klatt, 1973).

That the major correlate of perceived "contouricity" for ramps longer than 40 msec is the frequency range subtended is of interest in view of Pollack's (1968) experiments with frequency-modulated pure tones. Pollack found that the discrimination of brief FM tones ($\lesssim$27 msec) is governed by the size of the frequency excursion while for long transitions ($\gtrsim$400 msec) discriminability is proportional to the temporal differences subtended and

is nearly independent of the frequency range traversed. In between (27-400 msec), in the region most relevant to lexical tone, discrimination is a function of both the frequency range and temporal interval subtended. However, it is possible that Pollock's data is, at least partially, an artefact of the particular experimental paradigm used. In his study, ramps were bounded on both sides by a level frequency. Thus at brief transition durations, subjects may merely have compared the pitch of the initial level frequency with the pitch of the terminal frequency portion, the transition itself carrying no essential information about the pitch. The percept of moving pitch may thus require a minimum signal duration for the sensation of dynamic pitch to develop. More generally contour pitch may be based on the integration of information from successive "temporal quanta". That the discrimination of long frequency transitions is governed by the temporal difference subtended suggests the existence of a temporal quantum with limited memory. The reliance on both frequency range and temporal difference in the region between 27 and 400 msec might merely indicate that dynamic pitch is based on the integration of information from a minimum number of temporal quanta. Additional evidence for the confluence of two mechanisms - one based on frequency range, the other on temporal differences - comes from Klatt's (1973) study of F_o discrimination in synthetic speech in which he found that the difference limen is twice as great for F_o ramps with unequal rates of change as compared with ramps of identical slope (4 Hz vs 2 Hz).

III. EXPERIMENT II: THE CONTRIBUTION OF DURATION TO THE PERCEPTION OF CONTOUR TONES

The results of experiment I suggested that in order to perceive a fundamental frequency-modulated vowel as contour-like the ramp duration must exceed a threshold lying between 40 and 65 msec. However, it is possible that the critical parameter is not ramp duration per se but rather the duration of the entire signal (steady-state + ramp) or an interaction of the two. Experiment II explored this possibility by presenting for evaluation stimuli in which the duration of the initial steady-state F_o portion ranged from 0 to 160 msec while the ramp duration was held constant at 90 msec.

Stimuli. Stimuli were generated from the same five-formant vowel used in Experiment I. Stimulus duration ranged from 90 msec to 250 msec. The fundamental frequency of the briefest stimulus rose in approximately linear fashion (5 Hz increment per pitch period) from 100 to 150 Hz (Figure 4: solid line). Five other stimuli, each incorporating this same ascending ramp as its terminal portion were generated in such a manner that they differed from each other only with respect to the length of the initial steady-state portion. The steady-state duration varied in logarithmic steps from 10 to 160 msec (Figure 4: dotted line). The fundamental frequency of the steady portion was 100 Hz.

Method. Stimuli were presented over loudspeaker to 30 native English speaking members of an introductory phonetics class who were asked to rate the vowels on their degree of"contouricity" applying the same set of criteria listeners in Experiment I used. No individual served as a subject in both Experiments I and II. After a brief "warm-up" period consisting of two presentations of each stimulus, listeners rated 100 stimuli on a seven point scale. The first forty trials were discarded so that there were ten repetitions of each stimulus played in random sequence.

Results. There is a strong positive correlation between stimulus duration and degree of perceived "contouricity"(Figure 5). A one way analysis of variance indicated that the effect of increasing duration is both highly significant $F(5, 145) = 253.68$, $p<.0001$ and extremely linear (85% of the variance is accounted for by a linear component using the Newman-Keuls method of analysis). Individual T-Tests established that the "contouricity" of each stimulus was rated as being different from that of its closest neighbor ($p<.001$ for all pooled comparisons). That the results do not merely reflect a range effect in which ratings are based directly on stimulus duration is supported by a comparison of the rated degree of contouricity for a pair of similar stimuli, one used in Experiment I, the other in Experiment II. The stimuli are rated similarly (4.18 for a 90 msec ramp, incremented at 5 Hz per pitch period, and preceeded by an 80 msec level F_o in Experiment II. 4.48 for a 70 msec ramp incremented at 4 Hz per cycle and preceded by an 80 msec level F_o in Experiment I).

Discussion. The role played by signal duration in the processing of frequency-modulated sounds has been infrequently studied. Most of what is known about the interaction of pitch and duration has been the product of research on unmodulated (level) pure tones. For instance, a 1000 Hz. sinusoid loses all sensation of pitch if less than 10 msec long (Stevens and Davis, 1938; Dougherty and Garner, 1947). Tones briefer than 25 msecs are prone to pitch match errors (Dougherty and Garner, 1948). The discriminability of pure tones decreases markedly for durations under 100 msec (von Bekesy, 1929 [1960].) However, it is not clear precisely how these experimental data relate to the perception of frequency-modulated signals. One study of potential relevance is that of Nabelek, Nabelek, and Hirsh (1970) who found that a single pitch is associated with a frequency-modulated tone briefer than 40 msec regardless of the rate of frequency change. The results of Experiment I accord very well with this finding.

An observation made by subjects in a study of Brady, House, and Stevens (1961) may also be of relevance. Listeners considered a rapidly changing resonant frequency of 20 or 50 msec duration to have but a single pitch. However the same FM resonance was described as a "chirp" when followed and preceded by steady-state resonances.

IV. CONCLUSION

The perception of pitch does not bear a straightforward relationship to the physical signal. As with the perception of unmodulated frequencies, dynamic pitch is, in part, influenced by such factors as intensity and duration. However, the degree of disparity between the physical signal and resultant perception appears to be considerably greater with changing pitch.

The auditory system does not have the capacity to track and record every aspect of a dynamic signal. Instead, it seems to rely on particular frequency "landmarks" within the signal. Thus the initial and final portions of a glide are often the key elements determining the perceived pitch pattern.

However, even this process of sensory truncation is subject to certain temporal constraints. Frequency glides shorter than 30-40 msec are not perceived as glides. At most, for really steep ramps, their dynamic quality is perceptually minimal (Experiment I; also Nabelek, Nabelek, and Hirsh, 1970; Pollack, 1968).

For glides exceeding this temporal interval, the dynamic perception is probably governed by the frequency range subtended (Experiment I; Pollack, 1968). At durations beyond those used in the current series of experiments (≥400 msec), the factors governing the perceptual qualities associated with a frequency-modulated signal appear to change once more to incorporate rate of change as a primary element (Pollack, 1968). It is at this point that the behavior of the system changes from one characterized by a two point memory to one which is capable of evaluating the dynamic properties of the signal with greater precision. Pollack's (1968) study is significant for an additional reason as well. His data show that there are probably more than two distinct forms of dynamic pitch perception. His data implies the presence of a "temporal quantum" model for dynamic pitch in which signals whose duration do not span more than a single quantum will not be processed on the basis of <u>rate</u> of frequency change.

One prediction made by the temporal quantum model is that the percept of dynamic pitch should be sensitive to any signal manipulations that increase the duration of the signal as a whole. The results of experiment II are consistent with this prediction. However the results do not provide any clue as to the duration of the quantum.

The results of Experiments I and II together suggest a perceptual basis for the absence of a dynamic quality for the Chinese short tones. The duration of the frequency ramp of a contour tone must exceed 40 - 65 msec for the percept of "contouricity" to develop (Experiment I). However, even the degree of contouricity associated with a 90 msec ramp with no steady state precursor is small. It requires an initial steady state F_0 portion of 40 msec or greater to produce a substantial percept of dynamic pitch. Therefore the minimum signal duration required for contouricity is 130 msec - longer than the 120 msec limit on the length of the Chinese short tones.

Acknowledgements. The authors would like to express their appreciation to Louis Goldstein,Peter Ladefoged,and Jim Wright for their comments on an earlier version of this paper.

REFERENCES

Bekesy,G. von (1929) Zur Theorie des Horens;Uber die eben merkbare Amplituden- und Frequenzanderung eines Tones;Die Theorie der Schwebungen. Physik Zeits. 30:721-745. Translated as "Just noticeable differences of amplitude and frequency", in Experiments in Hearing. New York:McGraw-Hill,1960.pp. 207-238.

Brady,P.,House,A.,and Stevens,K. (1961) Perception of sounds characterized by a rapidly changing resonant frequency. Journal of the Acoustical Society of America 33:1357-1362.

Collins,M.J. and Cullen,J.K. (1978) Temporal integration of tone glides. Journal of the Acoustical Society of America 63:469-473.

Dougherty,J.M. and Garner (1947) Pitch characteristics of short tones. I. Two kinds of pitch threshold. Journal of Experimental Psychology 37:351-365.

Dougherty,J.M. and Garner,W. (1948) Pitch characteristics of short tones. II. Pitch as a function of tonal duration. Journal of Experimental Psychology 38:478-494.

Ewan,W.G. (1976) Laryngeal behavior in speech. Unpublished PhD thesis,University of California,Berkeley.

Fodor,J.,Bever,T. and Garret,M. (1974) The Psychology of Language:An Introduction to Psycholinguistics and Generative Grammer. New York:McGraw-Hill.

Fry,D. (1955) Duration and intensity as physical correlates of linguistic stress. Journal of the Acoustical Society of America 27:765-768.

Haggard,M.,Ambler,S.,and Callow,M. (1970) Pitch as a voicing cue. Journal of the Acoustical Society of America 47:613-

Klatt,D. (1973) Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception. Journal of the Acoustical Society of America 53:8-16.

Ladefoged,P. (1975) A Course in Phonetics. New York:Harcourt,Brace and Jovanovich.

Lieberman,P. (1965) Language,Intonation,and Perception. Cambridge:MIT Press.

Nabelek,I.,Nabelek,A.,and Hirsh,I. (1970) Pitch of tone bursts of changing frequency. Journal of the Acoustical Society of America 48:536-553.

Pollack,I. (1968) Detection of rate of change of auditory frequency. Journal of Experimental Psychology 77:535-541.

Stevens,S.S. and Davis,H. (1938) Hearing:Its Psychology and Physiology. New York:Wiley.

Sundberg,J. (1973) Data on maximum speed of pitch changes. STL-QPSR 4:39-47.

Wang,W.S-Y. (1967) Phonological features of tones. International Journal of American Linguistics 33:93-105.

Wang, W.S-Y. (1972) The many uses of F_0. In Papers in Linguistics and Phonetics. Dedicated to the Memory of Pierre Delattre (A. Valdman, ed.) The Hague: Mouton.

Zee, E. and Greenberg, S. (1977) On the perception of contour tones. Journ. of the Acoust. Soc. of America 62:S47.

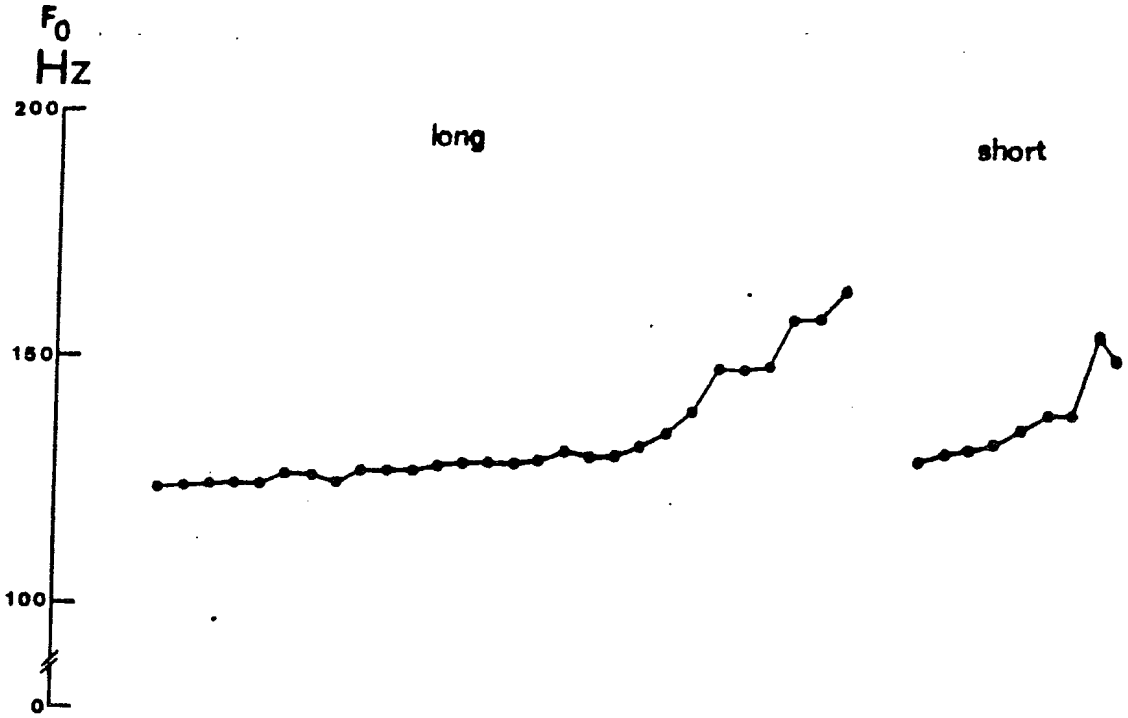


Figure 1. Digital pitch extraction analysis of two tonal classes in the Shanghai dialect of Chinese. Each dot represents the period analysis for a 25 msec speech sample. Dots are spaced at 10 msec intervals.

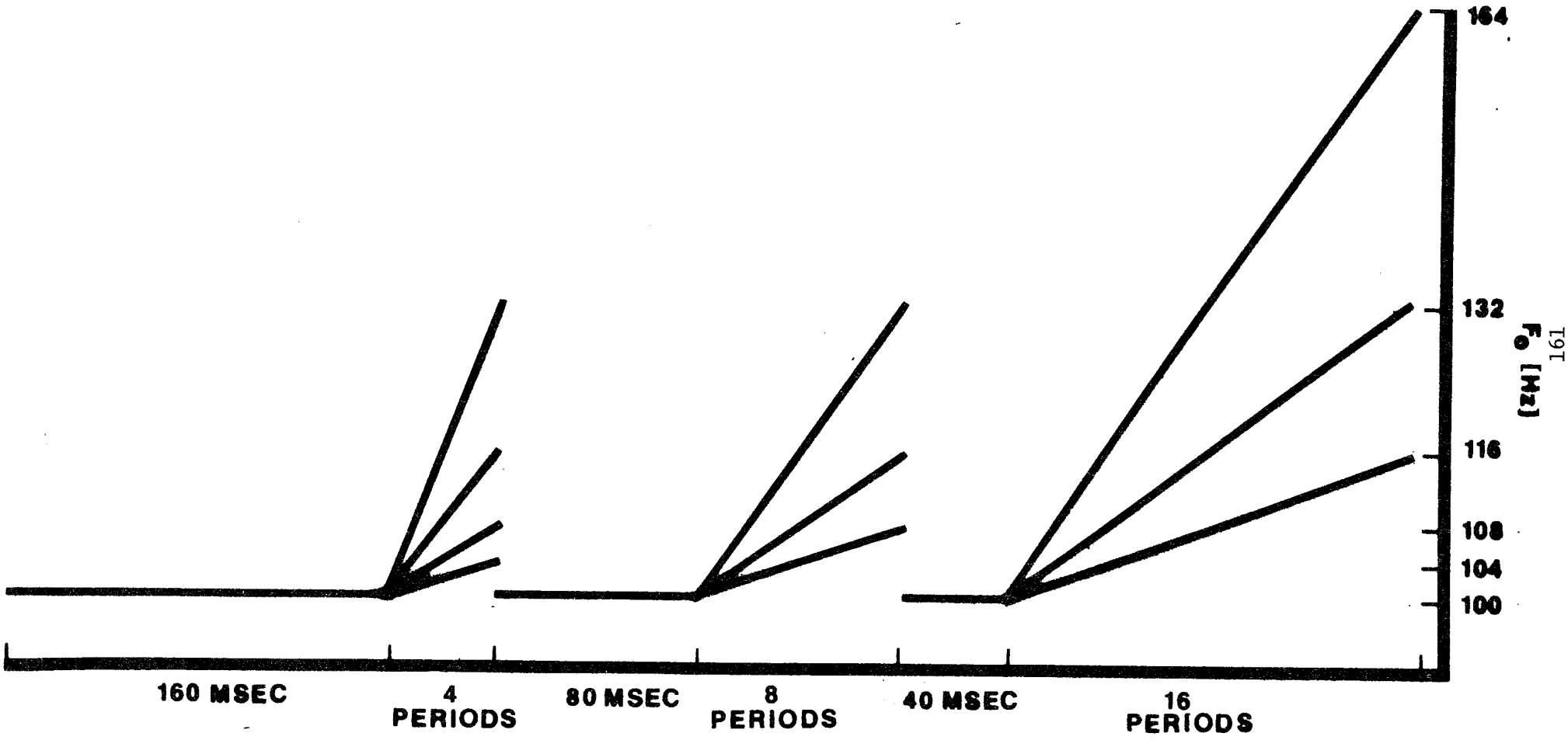


Figure 2. Fundamental frequency contours of stimuli used in Experiment I. The initial portion of each vowel had a steady F_o which varied in duration from 40 - 160 msec. The ascending terminal portion varied in duration from 33.5 - 39 msec for a 4 period ramp, from 68.2 - 76.5 msec for an 8 period ramp, and from 121.7 - 147.6 msec for a 16 period ramp.

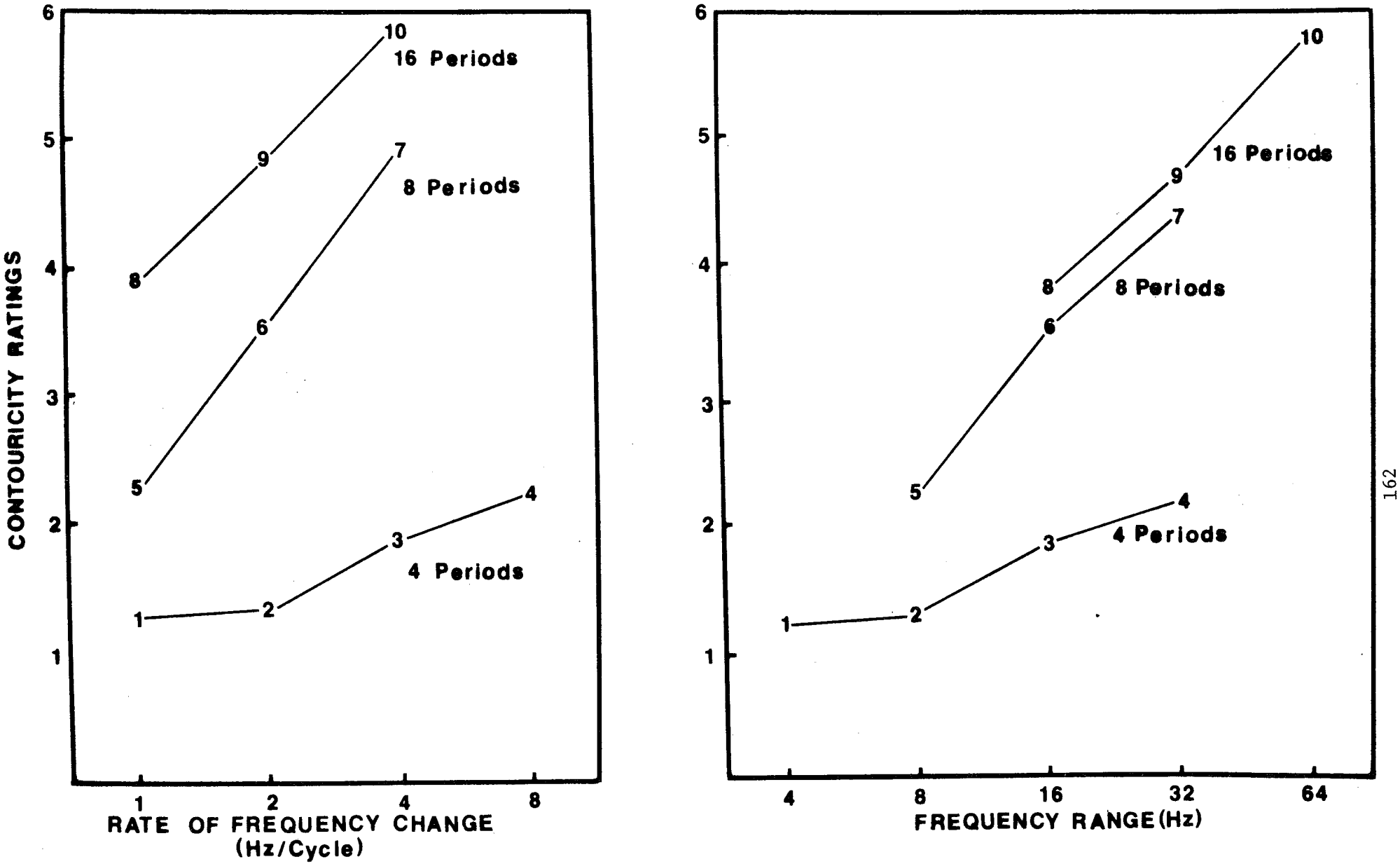


Figure 3. Degree of perceived "contouricity" as a function of (a) rate of change in fundamental frequency, and (b) range of fundamental frequency subtended. Digits connected by solid lines refer to stimuli specified in Table I.

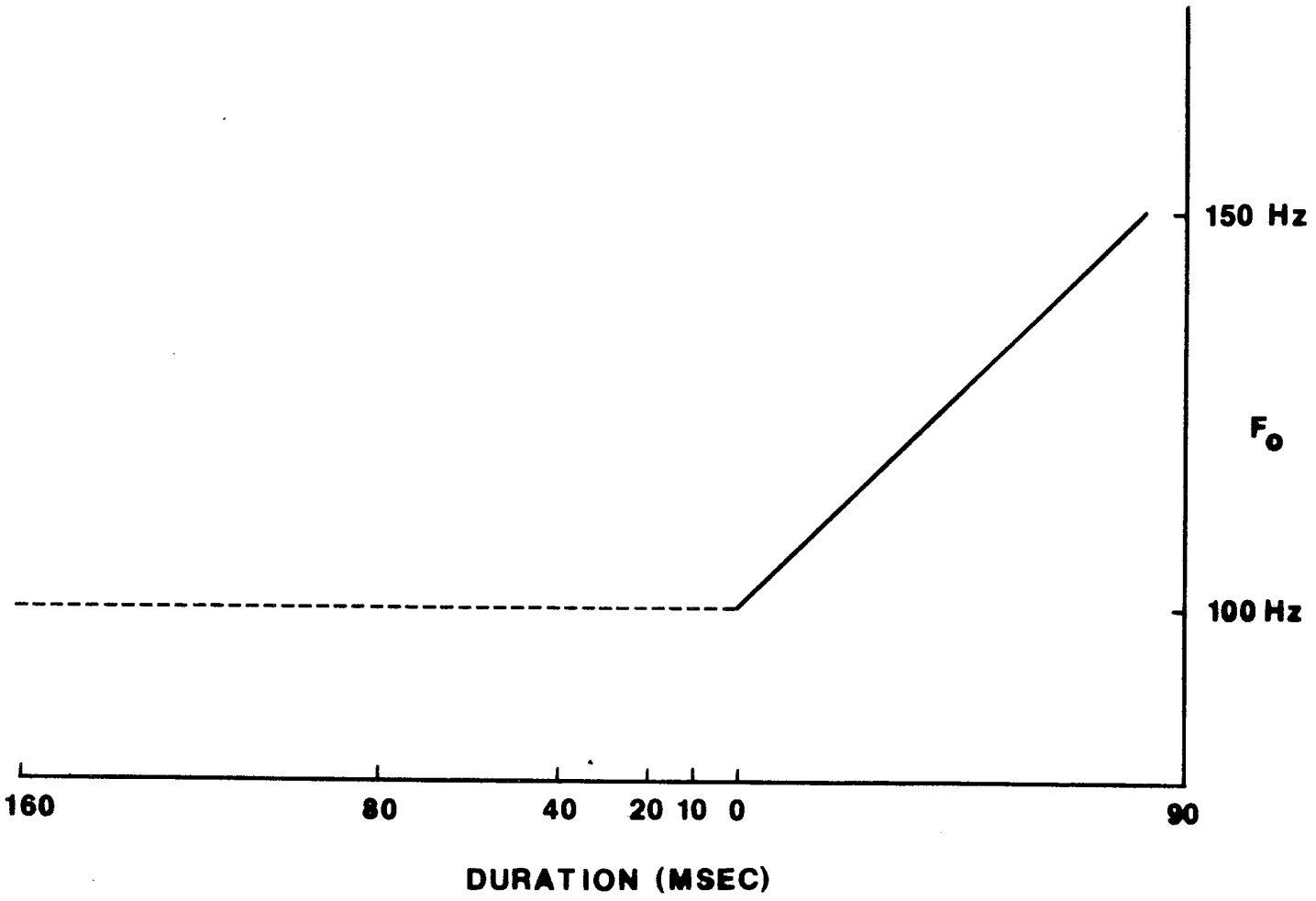


Figure 4. Fundamental frequency contours for stimuli used in Experiment II. The solid line represents an ascending F_o cntour which was present in all stimuli. Steady F_o segments of variable duration (0 - 160 msec) were adjoined to the ramp.

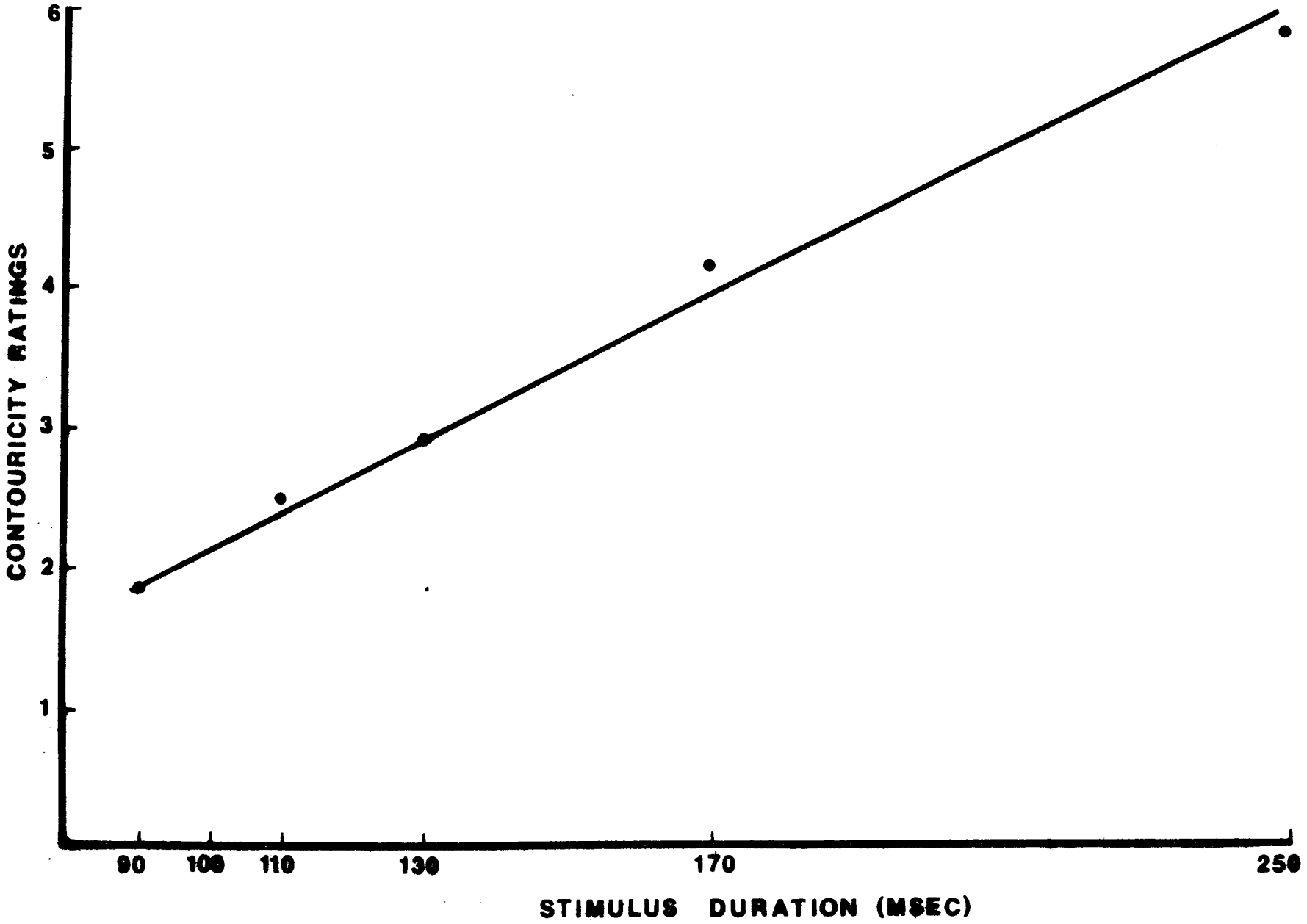


Figure 5. Degree of perceived "contouricity" as a function of overall stimulus duration. Duration is sum of steady state F_o segment and ramp segment.

Index of publications by members of the UCLA Phonetics Laboratory

October 1975 -- September 1978

Compiled by Jonas N. A. Nartey

Berinstein, Ava

1978. "A Cross Linguistic Study on the Perception of Stress." Journal of the Acoustical Society of America. 63: S-55 (Abstract).

"Acoustic Correlates of Stress in K'ekchi." Journal of the Acoustical Society of America. 64: S-93 (Abstract).

Browman, Cathe

1976. (with L. Goldstein) "Slips of the Ear - Misperceptions as Clues for Processing." Journal of the Acoustical Society of America. 59: S-56 (Abstract).

"Frigidity of Feature Detectors - Slips of the Ear." UCLA Working Papers in Phonetics 31: 62-67.

"The Natural Mnemopath: or, what you know about words you forget." UCLA Working Papers in Phonetics 31: 68-71.

1978. "Tip of the Tongue and Slip of the Ear: A Comparative Study." Journal of the Acoustical Society of America. 64: S-93. (Abstract).

"Tips of the Tongue and Slips of the Ear: Implications for Language Processing. UCLA Working Papers in Phonetics 42.

Carlson, Ronald

1976. (With N. McKinney) "A Hybrid Multiple-Channel Pitch-Frequency Analysis System." UCLA Working Papers in Phonetics 33: 203-216.

Disner, Sandra F.

1976. (with P. Ladefoged) "Trill Seeking." Journal of the Acoustical Society of America 60: S-45 (Abstract)

1977. "Cross-Linguistic Survey of Vowel Quality." Journal of the Acoustical Society of America 62: S-49 (Abstract).

1978. Vowels in Germanic Languages. UCLA Working Papers in Phonetics 40.

"Normalization Across Languages." Journal of the Acoustical Society of America 64: S-21 (Abstract).

Fromkin, Victoria

1976 "On the Formal Nature of Language and Linguistic Theories." in Handbook of Perception III. (eds. E. Carterette and M. Friedman). New York: Academic Press. 3-23.

"A Note on Tone and the Abstractness Controversy." Studies in African Linguistics, Supplement 6: 47-62.

(with G. Silva and L. Galloway) "Computer Processing of 'Tips of the Slongue'. Journal of the Acoustical Society of America 59: S-56 (Abstract).

"The Wheres and Whyfores of Preach Seduction." UCLA Working Papers in Phonetics 31: 22-26.

"The Interference between Phonetics and Phonology." UCLA Working Papers in Phonetics 31: 104-107.

1977 "Obituary for Harry Hoijer." Language 53.1: 169-173.

1977 "Some Questions Regarding Phonetics and Phonetic Representations." Linguistic Studies in Honor of Joseph Greenberg. (ed. A. Juillard) Anma Libri and Co., 365-380.

"Putting the emPHAsis on the Wrong syLLAble." Studies in Stress and Accent. (ed. L. Hyman) Southern California Occasional Papers in Linguistics. 4: 15-26.

(with F. Heath, and G. Silva) "A Program for the analysis of speech error data." Siglash, 10.1,2: 15-29.

"The phonology of Akan revisited." Language and Linguistic Problems in Africa. (eds. P.F.A. Kotey and H. Der-Houssikian) Columbia, So. Carolina: Hornbeam Press. 143-154.

1978 "Working group on speech errors: slips of the tongue, ear, pen, and hand." UCLA Working Papers in Phonetics 41: 68-69.

Tone: A Linguistic Survey (editor) Academic Press, New York.

Gandour, Jack

1975 "On the representation of tone in Siamese." Studies in Tai linguistics in honor of William J. Gedney, ed. by J.G. Harris and J R. Chamberlain, 1970-195. Bangkok: Central Institute of English Language, Office of State Universities.

"The features of the larynx: n-ary or binary?" Phonetica 32.4: 241-253.

"Evidence from Lue for contour tone features." Pasaa 5.2:39-52.

Review of Tai phonetics and phonology, ed. by J.G. Harris and R.B. Noss. Pasaa 5.2: 131-142.

1976 "A glimpse of shamanism in southern Thailand. Journal of the Siam Society 64. 1: 97-103.

"A reanalysis of some phonological rules in Thai." Studies in Tai linguistics in honor of Fang-Kuei Li, ed. by T.W. Gething, J.G. Harris and P. Kullavanijaya, 47-61. Bangkok: Chulalongkorn University Press.

"Counterfeit Tones in the speech of Southern Thai bidialectals." UCLA Working Papers in Phonetics 33: 3-19.

"Aspects of Thai tone" UCLA Working Papers in Phonetics 33: 20-22.

(with I. Maddieson) "Measuring larynx height in standard Thai using the cricothyrometer" UCLA Working Papers in Phonetics 33: 160-190.

(see I. Maddieson and J. Gandour "Vowel length before aspirated consonants).

1977 "Counterfeit tones in the speech of Southern Thai bidialectals. Lingua 41. 125-143.

"On the interaction between tone and vowel length: evidence from Thai dialects" Phonetica 24: 54-65.

1978 (with R.A. Harshman) "Cross-language differences in tone perception: a multidimensional scaling investigation" Language and Speech 21: 1-33.

"The perception of tone." Tone: a linguistic survey, ed. by V.A. Fromkin. New York: Academic Press.

"Cross-language study of tone perception. Linguistic variation: models and methods, ed. by D. Sankoff, 139-147. New York: Academic Press.

(with V.A. Fromkin) "On the phonological representation of contour tones" Linguistics of the Tibeto-Burman Area 4: 73-74

Godinez, Jr. Manuel
1978 "A comparative study of some romance vowels." UCLA Working Papers in Phonetics 41: 3-19.

Goldstein, Louis

1975 James F. Lackner and Louis M. Goldstein. 'The psychological representation of speech sounds' Quarterly Journal of Experimental Psychology 27:173-185.

1976 "What we listen for." UCLA Working Papers in Phonetics 31: 72-77.

1977 (With M. van den Broecke) "Response bias and subjective estimation of consonant frequency." Journal of the Acoustical Society of America. 61: S-65 (Abstract).

1978 Three Studies in Speech Perception: Features, Relative Salience and Bias. UCLA Working Papers in Phonetics 39.

Greenberg, Steve
1978 "Digital simulations of vowel processing in the peripheral auditory system." Journal of the Acoustical Society of America. 64:S-137 (Abstract).

(with J. D. Sapir) Acoustic correlates of 'big' and 'thin' in Kujamutay. Proc. of the 4th Annual Meet, Berkeley Ling. Society. 293-310.

(with J. C. Smith, J. T. Marsh and W. Borwn) Human auditory frequency-following responses to a missing fundamental. Science 201: 639-641.

Harshman, Richard
1976 (With H.J. Crawford and E. Hecht) "Marijuana intoxication effects on cognitive style and hemispheric dominance tested via lateral tasks." in Conference on Human Brain Function (eds. Walter, Rogers and Sinzi-Freed). L.A. Brain Info. Service. (Also in S. Cohen and R.C. Stillman, eds., The Therapeutic Potential of Marijuana. New York: Plenum).

(With R. Remington) "Sex, language and the brain, Part I. A review of the literature on adult sex difference in lateralization." UCLA Working Papers in Phonetics 31: 86-103.

(With R. Remington and S. Krashen) "Sex, language, and the brain: adult sex differences in lateralization." in Conference on Human Brain Function. (eds. Walter, Rogers and Sinzi-Freed). L.A. Brain Info. Service.

(With G. Papçun) "Vowel normalization by linear transformation of each speaker's acoustic space." Journal of the Acoustical Society of America 59: S-71.

1977 (With P. Ladefoged, and L. Goldstein) "Factor analysis of tongue shapes." in Journal of the Acoustical Society of America 62.3: 693-707.

Hombert, Jean-Marie

1975 "The perception of contour tones." Proceedings of the First Annual Meeting of the Berkeley Linguistic Society. 227-232.

"Noun classes of tones in Ngie." in Studies in Bantu Tonology. (ed. L. Hyman) Southern California Occasional Papers in Linguistics. 3: 3-21.

"Phonetic motivations for the development of tones from postvocalic [h] and [?]: evidence from contour tone perception." Report of the Phonology Laboratory. UC Berkeley 1: 39-47.

1976 "The effect of vowel quality on pitch." Proceedings of the 14th Congress of Acoustics. Bratislara. 40-42.

(With J. Ohala and E. Ewan) "Tonogenesis: theories and queries." Report of the Phonology Laboratory. UC Berkeley 1: 48-77.

"Phonetic explanation of the development of tones from prevocalic consonants." UCLA Working Papers in Phonetics 33: 23-39.

"Word games: some implications for analysis of tone and other phonological processes." UCLA Working Papers in Phonetics 33: 67-80.

(With S. Greenberg) "Contextual factors influencing tone discrimination." UCLA Working Papers in Phonetics 33: 81-89.

(See E. Zee "Intensity and duration as correlates of Fo.").

1977 "Development of tones from vowel height." Journal of Phonetics 5: 9-16.

"Consonant types, vowel height and tone in Yoruba." Studies in African Linguistics 8.2: 173-190.

"Perception of tones of bisyllabic nouns in Yoruba." Studies in African Linguistics. Supplement 6: 109-121.

"The difficulty of producing different Fo in speech. UCLA Working Papers in Phonetics-36: 12-19.

A model of tone systems." UCLA Working Papers in Phonetics 36: 20-32.

(With P. Ladefoged) "The effect of aspiration on the fundamental frequency of the following vowel." UCLA Working Papers in Phonetics. 36: 33-40.

"Tone space and universals of tone systems." Journal of the Acoustical Society of America 61: S-80 (Abstract).

1978 "Why are nonperipheral vowels avoided?" Journal of the Acoustical Society of America. 64: S-18 (Abstract).

Jacobson, Leon
1977 "Phonetic aspects of Dholuo vowels" Studies in African Linguistics, Supp. 7:127-136.

Janda, R. D.
1978 "Spectrographic evidence for uniformity of voicing in obstruent clusters." Journal of the Acoustical Society of America. 64: S-92 (Abstract).

Ladefoged, Peter
1976 "How to put one person's tongue inside another person's mouth." Journal of the Acoustical Society of America. 60: S-77 (Abstract).

(With Iris Kameny and W.A. Blackenridge) "Acoustic effects of style of speech." Journal of the Acoustical Society of America 59.1: 228-231.

(With P. Macneilage) "The production of speech and language." Handbook of Perception VII: Language and Speech. (eds. E.C. Carterette and M. P. Friedman) New York: Academic Press. 75-120.

(With K. Williamson, B. Elugbe, and A.A. Uwalaka) "The stops of Owerri Igbo." Studies in African Linguistics. Supplement 6: 147-163.

"The phonetic specification of the languages of the world." UCLA Working Papers in Phonetics 31: 3-21.

1977 "The abyss between phonetics and phonology" Proceedings of the Chicago Linguistic Society 13: 225-235.

(With L. Rice) "Formant frequencies corresponding to different vocal tract shapes." Journal of the Acoustical Society of America 61: S-32 (Abstract).

(With R. Harshman, L. Goldstein and L. Rice) "Vowel articulation and formant frequencies." UCLA Working Papers in Phonetics 38: 16-40.

"Some notes on recent phonetic fieldwork." UCLA Working Papers in Phonetics 38: 14-15.

1978 "Phonetic differences within and between languages." UCLA Working Papers in Phonetics 41: 32-40.

"Expectation affects identification by listening." UCLA Working Papers in Phonetics 41: 41-42.

(With M. Lindau) "Prediction of vocal tract shapes in utterances." Journal of the Acoustical Society of America. 64:S-41 (Abstract).

La Velle, Carl R.
1976 "Universal rules of tone realization." UCLA Working papers in Phonetics 33: 99-108.

Lindau, Mona

1976 "Larynx height in Kwa." UCLA Working Papers in Phonetics 31: 53-61.

(With P. Wood and P. Lafage) "Vowel spaces in two Kwa languages." Journal of the Acoustical Society of America 60: S-44 (Abstract).

1977 "Vowel features." UCLA Working Papers in Phonetics 38: 49-81.

(With P. Wood) "Acoustic vowel spaces." UCLA Working Papers in Phonetics 38: 41-48.

Linker, Wendy
1977 "Lip positions and formant frequency in American English vowels." UCLA Working Papers in Phonetics. 41: 20-25.

Maddieson, Ian
1976 "A further note on tone and consonants." UCLA Working Papers in Phonetics 31: 47-52.

"The intrinsic pitch of vowels and tones in Foochow." UCLA Working Papers in Phonetics 33: 191-202.

Tone reversal in Ciluba - a new theory." Studies in Bantu Tonology (ed. L. Hyman) Southern California Occasional Papers in Linguistics 3: 141-165.

1977 (With J. Gandour) "Vowel length before aspirated consonants." Indian Linguistics 38: 6-11.

"Tone loans: A question concerning tone spacing and a method of answering it." UCLA Working Papers in Phonetics 36: 49-83.

"Tone spreading and perception". UCLA Working Papers in Phonetics 36: 84-90.

"Further studies on vowel length before aspirated consonants." UCLA Working Papers in Phonetics 38: 82-90.

1978 "The frequency of tones." Proceedings of the 4th Annual Meeting of the Berkeley Linguistic Society. 360-369.

"Universals of tone." Universals of Human Language Vol. 2. Phonology. (eds. J.H. Greenberg, C. Ferguson and E. Moravcsik). Stanford University Press. 335-366.

Meyers, Laura

1976 Aspects of Hausa Tone. UCLA Working Papers in Phonetics 32.

Moskowitz, Breyne A.

1975 "The acquisition of fricatives: A study in phonetics and phonology." Journal of Phonetics 3: 141-150.

Nartey, Jonas N. A.

1978 "Discrimination perception of nasal and non-nasal vowels: a reaction time test." Journal of the Acoustical Society of America. 64: S-19 (Abstract).

(With I.D. Condax) "The epiglottis in speech. Journal of the Acoustical Society of America . 64: S-91 (Abstract).

Papcun, George

1976 "How may vowel systems differ?" UCLA Working Papers in Phonetics 31: 38-46.

(With P. Ladefoged) "Two 'voiceprint'cases." UCLA Working Papers in Phonetics 31:108-113.

(With R. Harshman) "How do different speakers say the same vowels?" Journal of the Acoustical Society of America 59: S-71 (Abstract).

1977 "Relationships among formant frequency measures of vowels in an imitation dialect." Journal of the Acoustical Society of America 61: S-89 (Abstract).

1978 "Discriminant analysis on an imitation dialect". Journal of the Acoustical Society of America 64: S-21 (Abstract).

Rice, Lloyd

1976 "Friends, humans and country robots: lend me your ears" BYTE 12.

"Hardware and software for speech synthesis." Dr. Dobbs Journal of Computer Calisthenics and Orthodontia. 1.4.

(With P. Ladefoged) "Uniqueness of articulations determined from acoustic data." Journal of the Acoustical Society of America 59: 59: S-70. (Abstract).

"Articulatory tracking of the acoustic speech signal." UCLA Working Papers in Phonetics 31: 32-37.

"A better LASS." Journal of the Acoustical Society of America 60: S-78 (Abstract).

1977 "Speech synthesis by a set of rules" Proceedings of the First West Coast Computer Faire. San Francisco.

Shayne, JoAnne
1976 (With S.M. Gass) "An investigation of the role of stress as a factor in speech perception." UCLA Working Papers in Phonetics 31: 78-35.

Terbeek, Dale
1977 A cross-linguistic multidimensional scaling study of vowel perception. UCLA Working Papers in Phonetics 37.

Van Lancker, Diana
1975 'Heterogeneity in language and speech: neurolinguistic studies' *Working Papers in Phonetics* 29.

1976 Neurolinguistics Bibliography. UCLA Working Papers in Phonetics 34.

(With V. Fromkin) "Cerebral dominance for pitch contrasts in tone language speakers and in musically trained and untrained English speakers." Working Papers in Phonetics 36: 41-48.

Zee, Eric
1976 (With J.M. Hombert) "Intensity and duration as correlates of Fo." Journal of the Acoustical Society of America 60: S-44

1977 (With S. Greenberg) "On the perception of contour tones." Journal of the Acoustical Society of America 62: S-47 (Abstract).

"Duration and intensity as correlates of Fo." UCLA Working Papers in Phonetics 36: 111-121.

"The effects of Fo on the duration of [s]." UCLA Working Papers in Phonetics 36: 122-127.

1978 "Tone and vowel quality." UCLA Working Papers in Phonetics 41: 53-61.

"Effect of vowel quality on perception of nasals in noise." Journal of the Acoustical Society of America 64: S-19 (Abstract).