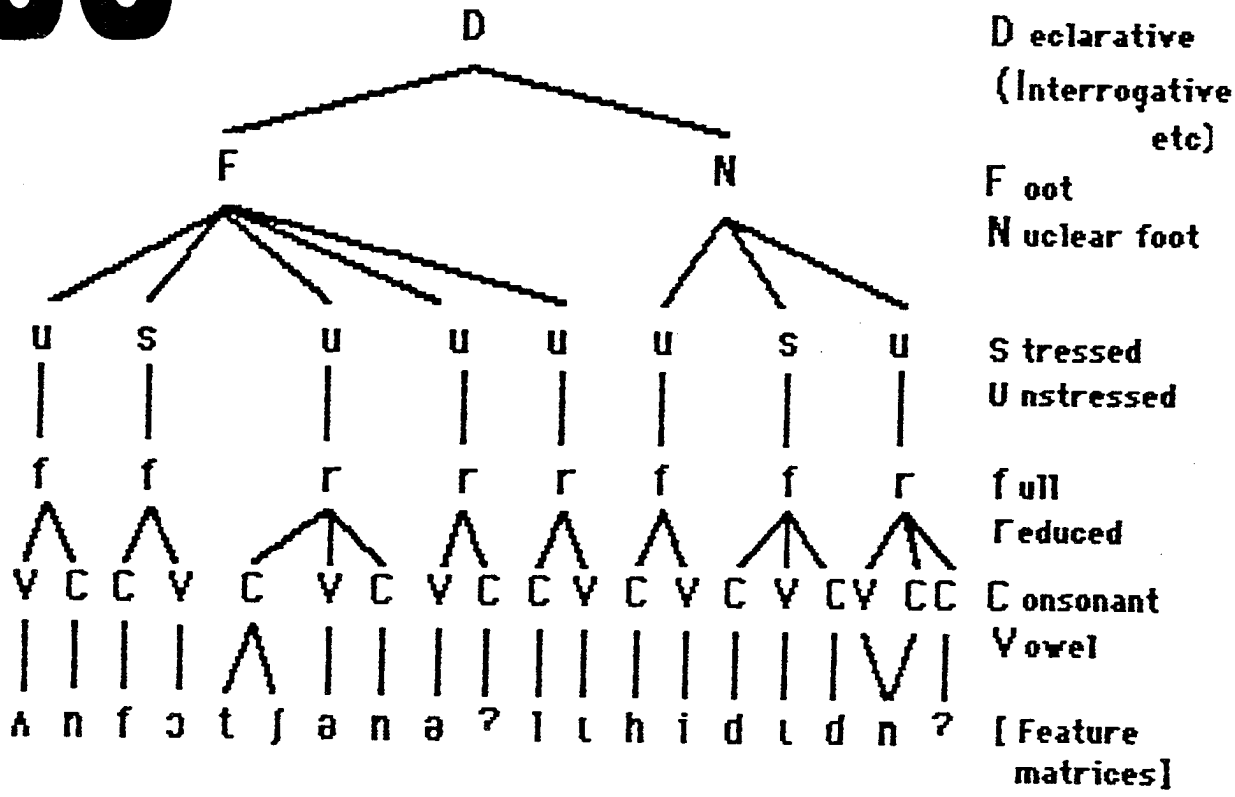
ucla working papers in phonetics

60



FEBRUARY 1985

The UCLA Phonetics Laboratory Group.

Stephen R. Anderson
Alice Anderton
Norma Antoñanzas-Barroso
Abby Cohn
Sarah Dart
Bill Dolan
Karen Emmorey
Vicki Fromkin
Bruce Hayes
Marie Huffman
Michel Jackson
Hector Javkin
Pat Keating
Paul Kirk
Jody Kreiman
Peter Ladefoged
Jenny Ladefoged
Karen Lau
Mona Lindau
Ian Maddieson
Carl Oberg
Alec J.C. Pongweni
Kristin Precoda
Ren Hong-Mo
Lloyd Rice
Mika Spencer
Henry Teheranizadeh
Diana Van Lancker

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

Funds for the UCLA Phonetics Laboratory are provided through:

NSF grant BNS-23110
USPHS grant 1 R01 NS18163-03
and the UCLA Department of Linguistics.

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

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

UCLA Working Papers in Phonetics 60

February 1985

On the Naturalness of Stop Consonant Voicing[1]

John R. Westbury and Patricia A. Keating

1. Introduction

A long-recognized problem for linguistic theory has been to explain why certain sounds, sound oppositions, and sound sequences are statistically preferred over others among languages of the world. The formal theory of markedness, developed by Trubetzkoy and Jakobson in the early 1930's, and extended by Chomsky and Halle (1968), represents an attempt to deal with this problem. It is at least implicit in that theory that sounds are rare when (and because) they are marked, and common when (and because) they are not. Whether sounds are marked or unmarked depends -- in the latter version of the theory, particularly -- upon the 'intrinsic content' of acoustic and articulatory features which define them. There has, however, been no substantive attempt to show what it is about the content of particular features and feature combinations that causes them to be marked and others not.

In the last ten to fifteen years, the theory of markedness has been supplemented with -- some might argue, supplanted by -- an increasingly popular notion of linguistic naturalness, developed and discussed to varying degrees by Ohala (1974, 1983), Hooper (1976), Stampe (1979), Vennemann (1972), Schacter (1969), Schane (1972), and Lindblom (1978, 1983; cf., also, Liljencrants and Lindblom, 1972). Those in concert with this notion maintain that the sounds, sound systems, and sound sequences that are most common among the world's languages are those that are most natural -- natural because they are somehow easiest to articulate or perceive; because they represent physical constraints inherent to speech producing and perceiving systems; or, because they otherwise represent optimal tradeoffs between competing demands of perception and articulation.

Unfortunately, 'explanations' of typological generalizations based on naturalness have, as a rule, been no more satisfying that those based on markedness. This is particularly true in the phonological literature. The claim, for example, that speakers are more readily disposed to produce some sounds and sound sequences than others can be meaningful only if we know specifically how that is so. Thus, the notion of naturalness effectively presupposes well-developed models which specify (1) limiting properties of the production and perceiving mechanisms, thereby defining possible speech behaviors, and (2) general principles which prioritize that range of behaviors. However, among those whose work relates most directly to the notion -- with the exceptions of Lindblom and Ohala -- development of suitable models has been notably absent.

In this paper, we consider the general question of whether it is more 'natural' for stop consonants to be voiced or voiceless. According to the myoelastic-aerodynamic theory of phonation (van den Berg, 1958), the vocal folds will oscillate only when there exists an adequate pressure drop and airflow across them. During stops, no air exits the mouth or nose, so that this condition is not obviously met. That observation suggests that voiced stops might be more difficult to produce, and thereby less 'natural', than their voiceless counterparts (Ohala, 1983). And yet, a great many languages have voiced stops, at least in some phonetic environments. Indeed, if voiced stops are generally hard to produce, why do they exist at all? Why are they so prevalent? Why aren't they

generally unstable, both synchronically and diachronically? Simple questions such as these readily lead to an examination of the actual means by which voicing might be produced and maintained during a stop. In this paper we present a systematic approach to this question based on defining an explicit model of the articulatory mechanism, and then using that model to test the effects on voicing of a variety of articulatory conditions.

2. The Model

2.1. Basic approach

To a first approximation, the vocal tract consists of two soft-walled cavities, the lungs and mouth. They are separated from each other by a constriction formed by the vocal folds, and separated from the atmosphere by constrictions at the velopharyngeal port and/or mouth opening. Over the course of an utterance, the volumes of both cavities, dimensions of various constrictions, and the mechanical properties of the vocal tract walls and vocal folds themselves may be controlled voluntarily and independently, thereby producing the familiar low-frequency variations in air pressures and flows characteristic of speech.

It is possible to write a set of differential equations whose solutions will describe the response of this physical system to variations over time in its control elements, e.g. the dimensions, and thereby cross-sectional areas of glottal, oral, and velopharyngeal constrictions. These control inputs over time correspond to physiological interpretations of the familiar row-by-column feature matrix representation of an utterance. The outputs or responses of the system are also time functions which describe the consequences of any particular set of initial conditions and control functions, in terms of air pressures and flows which can be observed at various points along the vocal tract. Relevant details of our implementation of such a model, developed earlier by Rothenberg (1968), are avilable elsewhere (Westbury, 1983; Keating, 1984). Other models incorporating the same general approach toward understanding the breath-stream dynamics of speech, though differing in their complexity and implementation, have been described by Muller and Brown (1980), Flanagan et al. (1975), Ohala (1976), and Scully (1969).

2.2. Assumptions made in modeling voicing

Subsequent sections of this paper describe expectations which can be developed by using such a model to investigate when and to what extent stop voicing is likely to occur. These expectations depend foremost upon two major assumptions. The first of these is that voicing will occur 'spontaneously' whenever the states of the glottis and vocal folds are suitable for voicing, and there exists a sufficient pressure drop between the trachea and pharynx. The model does not have any direct representation of vocal cords per se: the glottal opening alone is represented. In effect, we assume the condition on the glottal state to hold whenever a (constant) cross-sectional area of the glottal slit is a fair approximation of the average glottal area during a vocalic period. Vocal fold tension is not represented in the model. Rather, we assume a particular pressure drop across the glottis that depends on tension, to be necessary for voicing initiation and maintenance. The model is then used to determine when voicing will occur, by calculating when the pressure drop across the glottis exceeds that threshold. In all experimental cases to be presented, our discussions of 'voicing' are to be understood to mean a sufficient pressure drop for voicing.

The second major assumption influencing expectations derived from the model is that the acoustic and physiological realization of an utterance depends heavily upon a well-defined interpretation of the notion ease of articulation. An utterance which consists underlyingly of a serially-ordered string of states, each with its own defining properties, must be input to the model as a set of control functions that vary over time. Each such control function may itself be segmented into a string of steady states and transitions. We define the easiest string of adjacent states to produce -- and thereby, the most 'natural', from an articulatory point of view -- to be the one in which the velocities of articulatory transitions, in each and all control functions, are least. Though there are many complications and considerations that would have to be taken into account by a truly general characterization of ease of articulation[2], this definition is sufficient for our interest in determining what happens in the easiest cases of all, namely, cases when very few states change between segments.

In spirit, this characterization of the notion ease of articulation is not new (cf. Ladefoged, 1982, for example). The advantage of such a characterization, of course, is that it can be used within the context of a suitably explicit model of the articulatory mechanism, and of the control functions which drive it, to determine a continuum of articulatory ease which associates cost with particular sequences of speech sounds. A second, important aspect of this characterization is that it emphasizes the role which phonetic context must play in determining whether a sound is 'easy' or 'hard' to produce. Sounds are generally not produced in isolation in natural languages. Rather, any given sound is customarily bounded, at least to one side, by other sounds whose properties are certain -- and are always shown -- to influence its own acoustic and articulatory manifestation. It is not implausible to assume that the degree of difficulty in producing a particular acoustic or articulatory state will depend as much upon the inherent difficulty in maintaining that state, as upon the difference between that state and temporally-adjacent ones. For that reason, the 'ease' of stop consonant voicing can only be ascertained by considering stops in a variety of phonetic contexts.

3. Experiments on position in utterance

3.1. The medial case

Consider now the following problem: Is it more likely for a stop to be voiced or voiceless in an articulatorily 'simple' vowel+stop+vowel string, where the initial and final vowels are identical, and the stop is chosen so that its articulation between the flanking vowels involves as few changes in control parameters as possible? It is generally assumed that in such a case, stops will most naturally be voiced due to their 'assimilation' to neighboring voiced vowels. In order to use the breath-stream dynamics model to calculate whether and how long the conditioins sufficient for voicing might exist during such a string, input functions are devised which, in effect, fix as constant throughout the string all but one of the model elements subject to voluntary control -- namely, the oral constriction. Specifically, the tissues surrounding the lungs are considered stretched, so that subglottal pressure derives entirely from their elastic recoil, allowing the thoracic musculature to be quiescent. Moreover, there are no muscularly-induced changes in supraglottal volume, or in the mechanical properties of tissues surrounding the lungs and mouth. Too, the velopharyngeal port remains fully occluded. Finally, the vocal folds are appropriately and constantly adducted and tensed for voicing. Over the entire VCV string, only cross-sectional area of the mouth opening is varied, as it must,

first to produce a constriction in the mouth and then to release it.

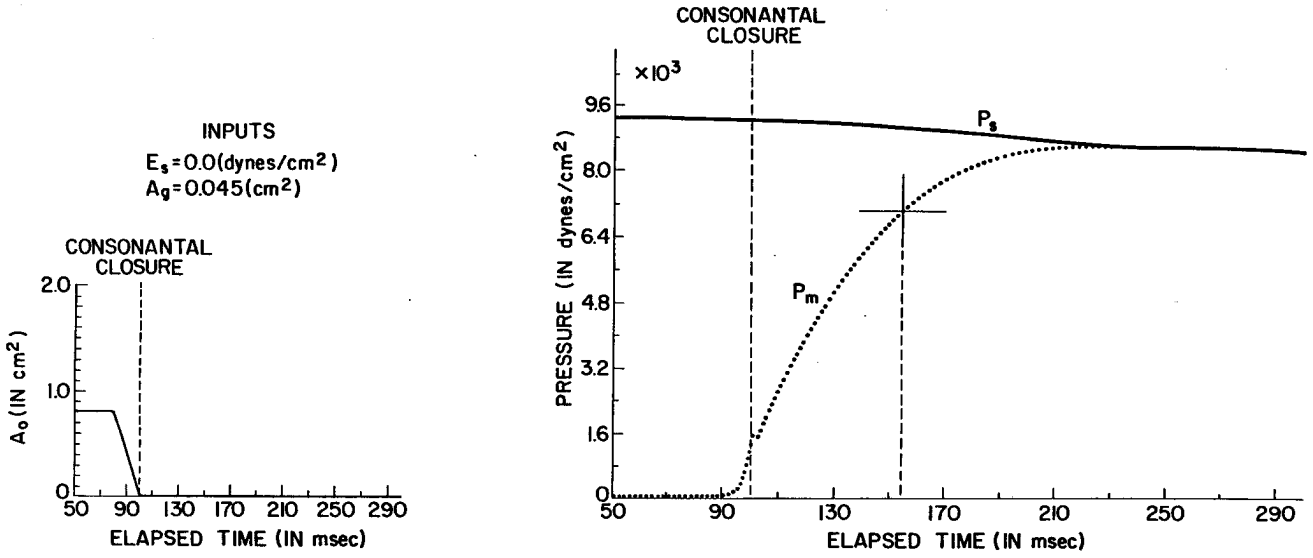


Figure 1. Calculated subglottal (P_s) and supraglottal (P_m) air pressure waveforms during an intervocalic labial stop consonant. The state of the vocal folds, suitable for voicing during the vocalic intervals preceding and following the occlusion, is assumed to remain constant during the consonant. Similarly, all other control elements in the model, except for that representing the oral constriction, have been set to constants.

Under conditions such as these, pressures above and below the glottis (P_m and P_s, respectively) can be expected to change with time as shown in Figure 1. There seems to be some consensus that the vocal folds will continue vibrating as long as the pressure drop across them is greater than roughly 2000 dyne/cm^2 (Ladefoged, 1964; Ishizaka and Matsudaira, 1972; Lindqvist, 1972; Baer, 1975). Note from this figure that the difference betwen P_s and P_m, though decreasing, is clearly greater than that amount for the first sixty-odd ms of the hypothetical 80 ms closure interval. Thus, voicing would be expected during that portion of the intervocalic stop, with offset occurring only late in the closure, within 20 ms of release for an 80 ms closure.

The relatively lengthy interval of closure voicing during such a stop like that simulated would be due almost entirely to the yielding walls which surround the supraglottal cavity. In effect, their outward motion during the stop closure -- in response to the increasing air pressure they contain -- retards the rate at which the transglottal pressure drop decreases, and thereby lengthens the interval during closure when voicing is possible. The precise duration of voicing will be influenced by factors which determine this expansion. The simulation illustrated in Figure 1 can be considered representative of a labial stop. For more posterior places of articulation, the extent of 'natural' closure voicing

would be somewhat less, since the total surface area surrounding the supraglottal cavity -- and thus, the total compliance of the supraglottal walls -- would also be less (Ohala and Riordan, 1979). Calculations with the model indicate that under otherwise similar conditions, the duration of closure voicing will be some 30% less for a velar stop than for a labial (Keating, 1983). Also, if the walls of the supraglottal cavity are assumed to be more lax than they are above (as well as in Figures 4 and 5) -- where they are assumed to have the compliance of tense cheek tissue (Ishizaka et al., 1975) -- the extent of 'natural' closure voicing may be greater. Calculations indicate that with such lax walls, voicing for all places of articulation will continue for at least 100 ms, that is, beyond the usual duration of singleton stop closures. On the other hand, if the walls are assumed to be rigid, effective pressure neutralization (and voice offset) will occur within 10 ms of occlusion. The same conclusion was reached by Rothenberg (1968). Stop voicing is difficult to maintain after a voiced sonorant only if the vocal tract walls are nearly rigid. We stress that rigid walls are not the usual case in speech production.

Lengthening the intervocalic closure, as might be appropriate in the simplest case for a homorganic stop cluster (or geminate) bounded by identical vowels, would have no effect on the expected time change in supraglottal pressure which occurs over the 80 ms closure interval depicted in Figure 1. Rather, lengthening the closure to something on the order of 150 ms would only allow that pressure more time to approximate pressure below the glottis. Thus, we might expect the closure of a relatively long, articulatorily-simple intervocalic stop -- in effect, a geminate or a homorganic cluster -- to be initially voiced and then voiceless. This simple conclusion is strongly reminiscent of an observation by Harms (1973) that the second /d/ in the phrase 'mad dog' often seems to be devoiced, though probably for 'natural' reasons, as Harms pointed out, rather than any intentional change in the glottal state. With an additional inference, one could approach Ohala (1983)'s conclusion that geminates are most naturally voiceless. If one assumes that stops must have more than half of their closures voiced to be perceived as voiced, then 60 ms of voicing out of 150 ms of closure would most likely be insufficient to establish a voiced percept, and therefore might result in the perception of geminates as voiceless. However, in pure articulatory terms, geminates appear to us to be most naturally partly voiced.

Of course, a medial stop (or homorganic cluster) may be fullly voiceless if articulatory adjustments occur at the level of the larynx, which make the vocal folds less susceptible to oscillation, and/or which hasten neutralization of the pressure drop across them. Tight adduction of the vocal folds, which often occurs in syllable-final voiceless stops in English (Fujimura and Sawashima, 1971; Westbury and Niimi, 1979), obviously has the former effect. Abduction of the vocal folds, which frequently accompanies the closures of voiceless stops in a variety of languages (Hirose, 1977), may have both effects. Alternatively, a medial stop (or cluster) may be made more fully voiced than the 60 ms or so suggested by Figure 1, by several methods, including contracting the expiratory muscles; decreasing average area of the glottis and/or tension of the vocal folds; decreasing the level of activity in muscles which underlie the walls of the supraglottal cavity; actively enlarging the volume of that cavity; or creating a narrow opening between the posterior pharyngeal wall and soft palate (Rothenberg, 1968; Westbury, 1983). These maneuvers, occurring singly or in combination, will have their greatest effect on duration of closure voicing when they occur during the closure inteval itself, in concert with the rise in

pressure above the glottis which naturally accompanies vocal tract occlusion. Implementing them in the model involves specifying how each of the relevant control parameters will vary in time.

However, within the context of the breath-stream dynamics model, additional glottal gestures leading to greater voicelessness, or other articulatory maneuvers leading to longer intervals of closure voicing, entail added cost in the sense that both involve changes in articulatory states above and beyond those (presumably) minimally necessary to produce the vowel+stop+vowel sequence. An assumption central to this paper is that changes in 'states' of articulators are the fundamental basis which determines articulatory cost or ease. If, indeed, speakers and languages seek out the easiest (ways to produce) sounds and sound sequences, then they should do so by minimizing changes in the various articulatory parameters subject to voluntary control. 'Speaking' in that fashion, using a simple model of the breath-stream control mechanism, suggests that a stop sandwiched between two identical vowels should naturally be largely voiced if its closure is relatively short, or voiced-voiceless if its closure is long.

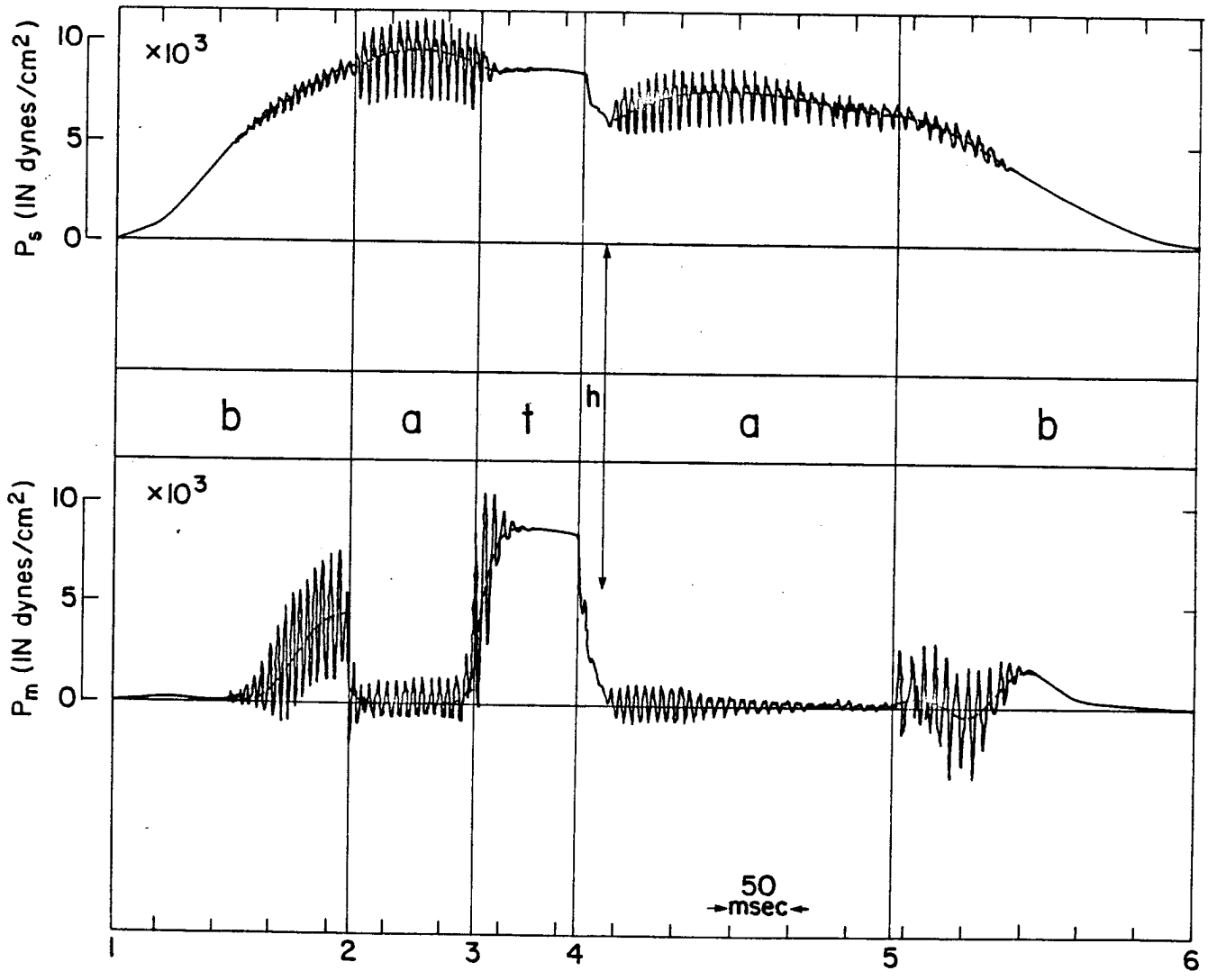


Figure 2. Subglottal (P_s) and supraglottal (P_m) air pressure waveforms recorded during the articulation of the nonsense disyllable /batab/ by one of the authors (JW). Oral closures are judged to occur at moments indicated by 1, 3, and 5, while releases are judged to occur at 2, 4, and perhaps 6.

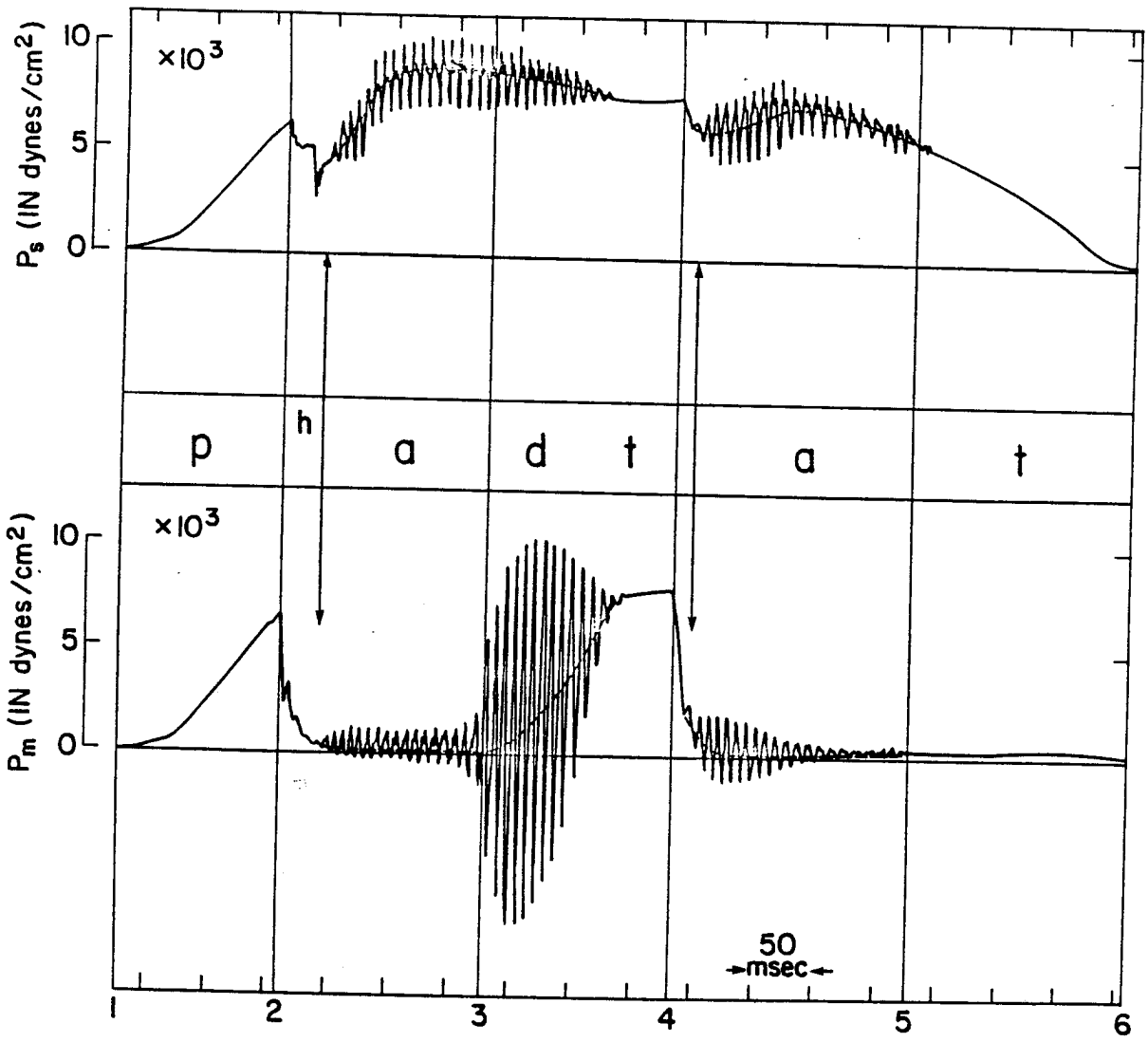


Figure 3. Subglottal (P_s) and supraglottal (P_m) air pressure waveforms recorded during the articulation of the nonsense disyllable /padtat/ by one of the authors (JW). Oral closures are judged to occur at moments indicated by 1, 3, ad 5, while releases are judged to occur at 2, 4, and perhaps 6.

3.2. The initial case

The expectations derived from the model and a simple characterization of ease of articulation regarding closure voicing during intervocalic stops and homorganic clusters are different from those regarding closure voicing during stops occurrng utterance-initially, or those occurring utterance-finally. One reason for these differences is straightforward. Note from Figures 2 and 3, for example, which show characteristic time functions of air pressures above and below the glottis during isolated disyllables produced by one of the authors (JW), that air pressure below the glottis remains generally high and stable over the middle segments of an utterance, but not over beginning (or ending) segments. Rather, in those respective environments, subglottal pressure rises above and falls toward the air pressure exerted by the atmosphere -- zero, in these figures -- in a generally linear fashion. Since the incidence of voicing depends in large

part upon the difference between subglottal and supraglottal pressures, stop voicing should be more likely utterance-medially than initially or finally, simply because that pressure difference in the former environment tends to be somewhat greater than in the latter environments.

Consider then in more detail the same problem for an utterance-initial stop considered previously for an intervocalic one: namely, is it more likely for a stop to be voiced or voiceless in an articulatory 'simple' string composed of pause + stop + vowel? As before, the model is used to calculate changes in air pressure during the string, determining the likelihood of closure voicing. However, specification of the articulatory conditions which affect the incidence and duration of closure voicing are necessarily different for hypothetical initial and intervocalic stops. As before, the input functions for an initial stop entail no muscularly-induced changes in supraglottal volume, or in the mechanical properties of tissues surrounding the lungs and mouth. Moreover, the velopharyngeal port remains fully occluded, and the vocal folds are appropriately and constantly adducted and tensed for voicing. However, cross-sectional area of the mouth opening is varied to release (rather than form) the oral constriction defining the stop. As before, the tissues surrounding the lungs are considered stretched, so that positive subglottal pressure will derive from their elastic recoil, but in the initial case the thoracic muscles are not considered quiet. Rather, the net force generated by the inspiratory muscles -- whose contractions initially enlarge the thoracic cavity -- is assumed to be slowly relaxing, in a roughly linear fashion over the closure interval of the initial stop. That slow relaxation 'checks' the recoil of the stretched thoracic tissues and can thereby provide a slow, smooth rise in subglottal pressure of the sort pointed out in Figures 2 and 3. Moreover, its inclusion among the input conditions for an utterance-initial stop is consistent with experimental observations made some years ago by Draper et al. (1959).

Given these input articulatory conditions, air pressures above and below the glottis can roughly be expected to vary with time, during an utterrance-initial stop, as shown in Figure 4. Subglottal pressure begins a steady rise roughly 200 ms before consonantal release, in a fashion similar to that illustrated previously in Figure 2. However, supraglottal pressure also rises steadily during closure, virtually in synchrony with subglottal pressure. That is because of our assumption that the 'voicing state' exists when the vocal folds are continuously apart, though narrowly so, separated by slightly more than .02 cm. As a consequence, the difference between subglottal and supraglottal pressures never exceeds the assumed voice-initiation threshold of 4000 dyne/cm^2 (cf. Baer, 1975), prior to the consonantal release. Instead, after the closure is released, air flows to the atmosphere, and intraoral pressure drops abruptly while subglottal pressure remains high. Only then is a suitable pressure drop achieved. Thus, the stop in an articulatorily simple pause + stop + vowel string will plausibly be voiceless (and unaspirated). Note that this result does not depend on having the glottis in either a voiceless or a breathing configuration. Either of those configurations would of course make voicing even less likely. However, if the vocal cords were fully adducted well before consonant release, such that no air passed between them, then a different pattern could result.[3]

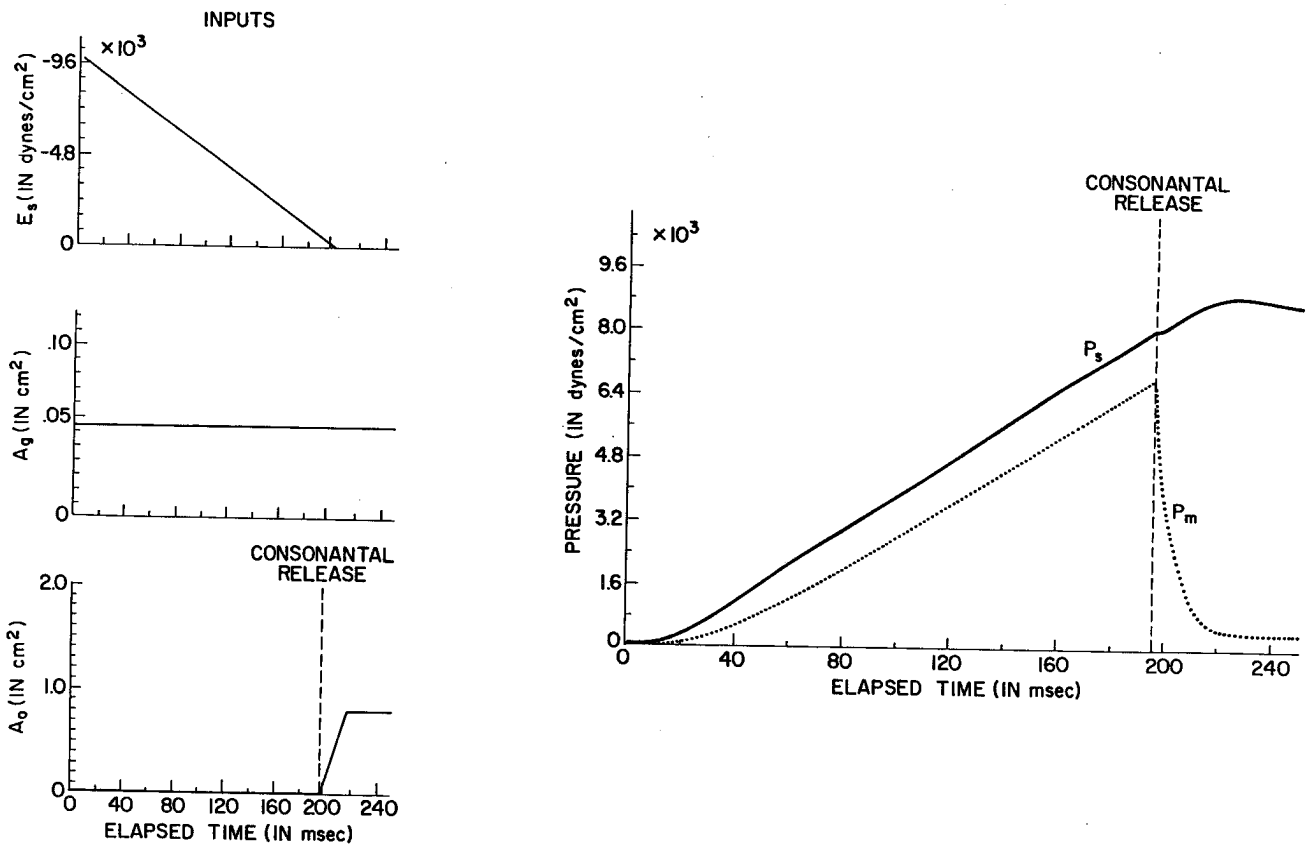


Figure 4. Calculated subglottal (P_s) and supraglottal (P_m) air pressure waveforms during an utterance-initial labial stop consonant. The steadily decreasing inspiratory force (E_s), represented here by convention in terms of a negative pressure head returning to zero, checks elastic recoil of the stretched tissues surrounding the subglottal cavity, thereby causing the slow, linear increase in subglottal pressure. The constant glottal area of 0.045 sq. cm is an approximation of the average area of the glottis during a sustained vowel. This figure suggests that voicing could begin only after release of the oral constriction (A_o) for an utterance-initial stop.

3.3. The final case

In an articulatorily-simple string composed of vowel + stop + pause, air pressures above and below the glottis can roughly be expected to vary with time as shown in Figure 5. All but two articulatory inputs for this simulation are the same as those for the intervocallic stop. Cross-sectional area of the mouth opening has been varied to produce a vocal-tract constriction, but (in this case) not to release it. Moreover, elastic recoil of stretched tissues surrounding the lungs is still thought to be responsible for the positive pressure head below the glottis, but that pressure is progressively being opposed by a hypothetically-increasing inspiratory force which begins 20 ms or so after the moment of oral occlusion. (The latter input to the model, admittedly ad hoc, provides in rather simple fashion a relatively linear decrease in subglottal pressure following implosion for a final stop of qualitatively the same sort that is apparent from either Figure 2 or 3.) Given these conditions, subglottal

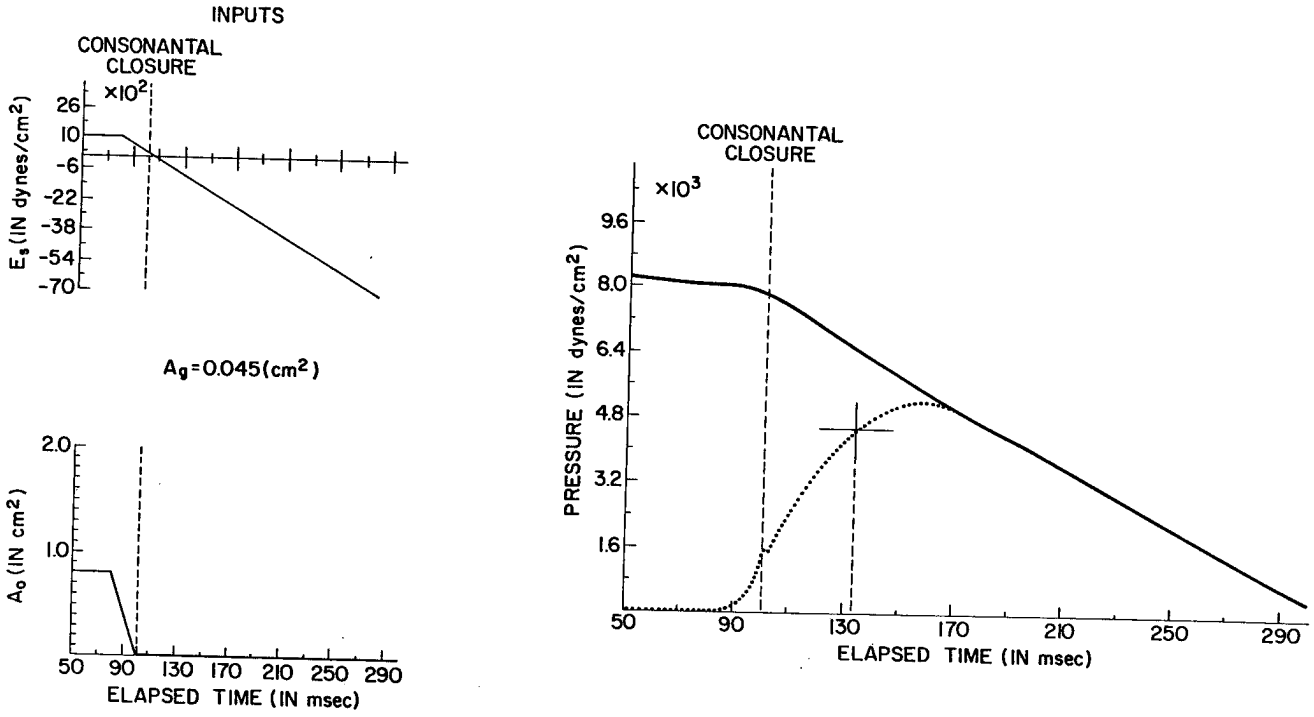


Figure 5. Calculated subglottal (P_s) and supraglottal (P_m) air pressure waveforms during an utterance-final labial stop consonant. The manipulation of respiratory forces (E_s) shown here has the articulatory interpretation of being initially expiratory, as might be expected during the latter phase of an utterance (cf. Draper et al., 1959), but quickly becoming inspiratory not long after closure for the final stop. Voice offset would likely occur near the cross intersecting the supraglottal air pressure waveform, circa 35 ms following closure of the oral port (A_o).

pressure can be expected to decrease after implosion for a final stop, in a roughly linear fashion analogous to the final decays in that pressure apparent in Figures 2 and 3. At the same time, intraoral pressure can be expected to increase. The two pressures rapidly converge and voicing offset occurs slightly more than 30 ms after the moment of occlusion. If the stop in an articulatorily-simple vowel + stop + pause string is held for something on the order of 100 ms, its closure will then be largely (though not entirely) voiceless. This result does not depend on a devoicing gesture in the glottis in preparation for respiration. Even with the vocal folds adducted, voicing will cease at some point during the stop closure.

Thus, to summarize these experimental results, consonant voicing will depend in part on position in utterance, to the extent that those positions differ in their subglottal pressure characteristics. Rising subglottal pressure in initial position and falling subglottal pressure in final position, if no counteracting steps are taken, both make voicing less natural than voicelessness. Relatively high and steady subglottal pressure in medial position makes voicing more natural than voicelessness. No differences in glottal configurations needed to be assumed to generate these differences; in all cases we assumed that the vocal cords were in a 'voicing state' conducive to vibration.

4. Comparison with language data

By hypothesis, voiceless initial and final stops, voiced intervocalic stops, and voiced-voiceless intervocalic stop clusters are relatively easier to produce than others. These easiest of stops are all articulated with the same 'voiced' configuration of the vocal folds. The extent to which voicing is actually realized during their closures depends then upon characteristic subglottal pressure functions associated with different positions in an utterance.

If languages seek out and prefer segments and utterances which are easiest to produce, and if the characterization of ease of articulation followed herein is fair, then we might expect to find that languages without phonemic voicing contrasts should maintain the articulatorily optimal stop system wherein context-dependent phonetic manifestation of the only 'type' of stop -- i.e., the natural, easy one -- would be variable but also easily predictable in terms of voicing. Moreover, we might expect that languages with phonemic voicing contrasts which also evidence neutralization of them in some environments should show a preference for the easier stops at neutralization sites. Neutralization to the optimal singletons and clusters would have the effect of making the utterances containing them easier to produce than they would otherwise have been. Conversely, maintaining any neutralized phonetic variant other than the optimal one in such an environment would entail added articulatory cost for the language.

Determining whether these expectations hold across various languages is difficult given information readily available in the literature. Certainly it is well-documented at the phonemic level that voiceless stops are preferred over those that are voiced. Ohala (1983:194) notes, for example, that of the '706 languages whose segment inventories were surveyed by Ruhlen (1975) 166 have only voiceless stops and 4 have only voiced stops'. Maddieson (1984) also finds support for this generalization from the UCLA Phonological Segment Inventory Database. The conventional account for this 'decisive "tilt" toward voicelessness' depends foremost upon the claim that 'there is a well-recognized difficulty in maintaining voicing during a stop' (Ohala 1983:194). However, the claim underlying this account is too strong. Under certain conditions, there is a well-defined sense in which it is more difficult to terminate than to allow (or maintain) voicing during a stop. Those conditions are not exotic. Rather, as we have shown, they can plausibly be expected quite frequently in at least one, very common phonetic environment, namely, utterance medial position. Thus, our expectations regarding stop voicing can be evaluated only by examining, in typological terms, patterns of allophonic variation within and among various languages.

Detailed phonetic data on common allophones -- particularly for languages without phonemic voicing contrasts, and to a lesser extent, for languages with contrasts which evidence morphophonemic alternation or surface neutralization of stop voicing -- are surprisingly hard to come by. The best source known to us is a recent review by Keating, Linker, & Huffman (1983) which provides a general description of allophonic variation among voiced and/or voiceless stops in 51 languages. However, even those data are less than ideal, first because they are largely categorial in nature, and secondly, because they are acoustically based (in the sense that they are derived from transcriptions by phoneticians and/or from spectrographic and oscillographic analysis) rather than physiological.

The problem with 'categorial' data becomes manifest, for example, in a conventional description of modern Polish where utterance-final instances of /b,d,g/ are said to be devoiced. Similarly, final voiced stops in the speech of children acquiring American English are frequently described as devoiced. However, oscillographic analyses have shown that the closures of final /b,d,,g/ in both Polish (Gianinni and Cinque, 1978) and the developing speech of young children (Smith, 1979) reveal more closure voicing (ca. 30-40 ms) than do their underlyingly voiceless counterparts (ca. 10-20 ms). Similarly, in Catalan the closure durations of devoiced underlying voiced stops differ from those of underlying voiceless stops (Dinnsen and Charles-Luce 1984). At least in acoustic terms, in each case the underlying voiced consonants are not identical on the surface to the underlying voiceless consonants. Rather -- and it is not an uninteresting fact -- they are merely labeled similarly. However, that act of categorial labeling -- inherent in a segmental inventory, conventional descriptive phonology, or phonetic transcription -- obscures acoustic and physiological details such as these which may be crucial for a fair test of our hypotheses.

Even noncategorial acoustical data are not ideal for our purpose of testing claims about articulatory ease. Only physiological data in terms comparable to our model are really adequate to that task. Thus, when we do find correspondences between language data and the model's predictions, we will not be sure that the predictions are borne out for the right reasons. Nonetheless, with this reservation in mind, we can at least ask when data from natural languages appear to conform to the expectations derived from the model. After all, if even acoustic correspondences are not found, then there is no point in entertaining the articulatory hypotheses we have presented.

Consider first the six languaes described in Keating et al. (1983) -- Alyawarra, Hawaiian, Kaititj, Nama, Tiwi, Yidin^{y} -- which exhibit no phonemic voicing-related contrast among stops. At least in acoustic terms, we can say that these languages -- as a group -- prefer voiceless stops in all permissible environments, and show less in the way of allophonic variation regarding stop voicing than languages which have phonemic contrasts. These languages create the impression that variability in the acoustic realization of the same speech sound in different environments is generally undesirable. In simple terms, our most preferable stop 'system' -- including voiceless unaspirated stops initially, voiced singletons (and voiced-voiceless clusters) medially, and largely-voiceless stops finally -- does not seem to occur. Since most of these languages do not allow final stops of any sort, the main difficulty for our proposal comes from medial stops which are voiceless and unaspirated, rather than voiced.

We may, however, draw some consolation from the fact that our optimal stop system does seem to be reflected in 'developing languages' without contrasts -- in the speech of young children who are acquiring their native tongue. Prior to their mastery of whatever contrasts may exist in the speech of adults, children tend to articulate utterance-initial stops which are most often described as voiceless and unaspirated (Preston, Stark, & Yeni-Komshian, 1967; Eilers, Oller, & Benito-Garcia, 1984). Similarly, prior to their mastery of final contrasts, children tend to produce 'voiceless' stops before pauses (Smith 1979). Lastly, children seem to show in their earliest speech a greater preference for voicing during utterance-medial stops than do adults (cf. Smith, 1973), though this last generalization is clouded by considerable individual variation and differences in the temporal control of stop closures (Smith, 1978).

Consider next the more than forty languages described in the Keating et al. (1983) summary which maintain some system of voicing-related contrasts among stops, in at least some environments. The prevalence of voicing contrasts indicates that voicing, compared to other possible laryngeal contrasts, is a fairly easy contrast for languages to use. We have indicated the relatively simple articulatory adjustments that can be made to generate contrasting voicing categories of consonants, relative to the kinds of adjustments that can be made for other kinds of contrasts. At the same time, we can consider those cases in these languages in which there is voicing neutralization, meaning either rules of synchronic alternations, or defective distributions. If, as we have hypothesized, languages prefer stops which are easiest to produce, then we might expect them to do so at least in environments where neutralization of an existing voicing contrast occurs. Maintaining any neutralized phonetic variant other than the optimal one in such an environment entails added articulatory cost for the language.

Generalizations from the Keating et al. (1983) review relevant to this expectation are as follows:

First, neutralizaton of a stop voicing contrast in utterance-initial position is uncommon. There are only four languages in the sample which, for at least some places of articulation, contrast voiced and voiceless stops medially and/or finally but not initially. These include Cuna, Efik, Ewondo, and Tamil. In all but the second of this group, the observed neutralization exhibits the expected 'destination' -- i.e., a voiceless unaspirated stop. In Efik, however, a phonemic contrast between labials in medial position is neutralized to a voiced labial utterance-initially.

Secondly, neutralization of voicing contrasts medially appears to be quite rare. In American English, underlying /t/ and /d/ neutralize to the voiced flap [D] in specific medial environments (Zue and Laferriere, 1979), though some contrast between them is maintained initially, finally, and in other medial environments. Similarly, in Zoque, an initial contrast between /tj/ and /dj/ is also neutralized medially, though contrary to our expectaton, to a voiceless allophone.

Thirdly, in languages which maintain voicing contasts of one sort or another among stops occurring initially and/or medially, final position is far and away the most common site for neutralization. Nineteen of the fifty-one languages surveyed by Keating et al. (1983) exhibit at least some neutralization of voicing-related contrasts among stops in that environment. Those include Basque, Bulgarian, Cantonese, Choctaw, Cuna, Danish, Dutch, Efik, Ewondo, Gaelic, German, Korean, Polish, Russian, Spanish, Thai, Tikar, Vietnamese, and Zoque. In fifteen of these, neutralization proceeds to preferred targets, but other, 'unexpected' results obtain in Dutch, German, Spanish, and Zoque.

Together, these generalizations show that the majority of known cases of contrast neutralization involving singleton stops seem compatible wih expectations derived from the model. It must be stressed, however, that in most cases the domain under consideration in the literature survey is the word, not the utterance. For example, devoicing of phonemically voiced stops may obtain in final position within a syllable, a word, or an utterance in various languages. The expectations from modeling do not account for positional effects other than those associated with position in utterance. Other effects would be language-specific generalizations of the utterance patterns. However, it should

be noted that some effects commonly reported as 'word' effects are in fact constrained by pause, that is, are utterance effects. For example, Polish has a rule of 'word-final' consonant devoicing that applies only before pause. Before a vowel initial word, devoicing will not apply. In fact, in some dialects voicing is truly utterance conditioned: not only will devoicing not apply, but word-final voiceless stops will voice before a vowel. This is precisely the situation our modeling leads us to expect.

The survey in Keating et al. (1983) does not provide any data regarding contrast neutralization in medial clusters. However, we feel that it can safely be said that assimilation of voicing in medial stop clusters -- yielding neutral allophones from two or more sources -- is thought to be quite common across languages. Certainly, our prediction regarding the most preferable medial cluster is not borne out by the few facts known to us. The optimal cluster -- which in acoustic terms, might be described as having fully dissimilated members, but which in physiological terms has fully assimilated members -- is clearly not the customary destination of voicing neutralization (or assimilation) rules. Languages which have medial clusters underlyingly typically either maintain all voicing-related contrasts among them at the surface (e.g., American English, Punjabi, Hindi), or partially collapse underlying contrasts to yield two distinct surface forms whose respective members are at least acoustically assimilated (e.g., Russian, Dutch, French, Hungarian). Thus, the number of surface contrasts among medial clusters is either the same as or slightly reduced relative to the number of underlying contrasts. But, as far as we know, no language reduces all underlying contrasts among clusters to a single surface form.

These facts suggest two things. Foremost, there must be considerable pressure not to collapse entirely underlying systems of contrast. Reducing the voicing contrasts among all underlying stop clusters to a single, maximally-simple surface form might indeed be preferable from a physiological point of view, but that reduction would at the same time be decidedly less than optimal from what we might call an information-transfer point of view. 'We speak to be heard in order to be understood' (Jakobson et al, 1963:13). The utility of contrasts for conveying information -- i.e., for making ourselves understood -- is obvious. Surely, then, the extent to which we allow articulatory principles to govern our linguistic behavior depends upon the extent to which they impinge on its primary function.

>par Secondly, the sheer prevalence of languages which exhibit voicing assimilation in medial clusters suggests to us -- following the generally intuitive line we have tried to maintain (and exploit) -- that acoustic voicing assimilation in medial clusters must be relatively easier, in some physiological sense, than certain types of acoustic dissimilation which depend upon parallel differentiation at a physiological (articulatory) level. That is, making a medial cluster fully voiced or fully voiceless must be somehow articulatorily easier than making one initially voiced and finally voiceless by some means other than the 'natural' method, or initially voiceless and finally voiced. One line of reasoning which maintains the ease-of-articulation principle introduced earlier in this paper, and which 'prefers' clusters of the former rather than the latter sort, is the following:

In general terms, assimilation can be thought of as a reduction in the rates of articulatory changes which are specified to occur between adjacent or temporally-proximal segmental states in the underlying representation of an utterance. At the phonetic surface, that reduction can be manifest as (1) a

slowing of articulatory transitions, so that they (or perhaps, the segmental 'steady' states themselves) spread out in time, and/or (2) a reduction in the differences between adjacent states, so that one or all are undershot. If -- for reasons unknown -- voicing must be acoustically and physically maintained during the latter portion of the lengthy closure of a medial stop cluster, we know from experience with our vocal tract model that a speaker must expend extra articulatory effort. That effort might take several articulatory forms, but if its acoustic effect is to be local only to the latter portion of a lengthy closure, it must be expended in a ballistic or 'step' fashion. By hypothesis, expenditures of the latter sort are articulatorily costly. Alternatively, a speaker might make qualitatively similar articulatory adjustments to maintain voicing cluster-finally, but begin them sooner, in effect sacrificing acoustic (and physiological) integrity of the cluster-initial stop in favor of easing articulation of the string which contains it.

Similar reasoning can be used to argue that regressive assimilation of voicelessness in underlyingly voiced-voiceless clusters might be preferable to maintaining acoustic dissimilarity between clusters' members. If voicelessness local to an underlying voiced-voiceless cluster's final portion is to be insured by vocal fold abduction, and if subsequent glottal adduction for a following sonorant must be largely complete before the cluster is released, then the sequence of stops comprising the cluster may be relatively easier to articulate if the changes in glottal state are slowed by 'spreading backward in time.' The acoustic consequence of that spread, of course, would be that the cluster-initial stop would appear less voiced than if no change in glottal state were to occur.

In general, then, if the closures of medial clusters must encompass certain articulatory maneuvers to insure voicing and voicelessness, initially or finally, and if their closure durations cannot be extended appreciably, then it may be easier to spread those maneuvers in time, 'causing' assimilation of voicing (or voicelessness) among the clusters' members, than to make the maneuvers rapidly enough to limit the temporal domain of their effects.

5. Summary and conclusions

How well does aerodynamic modeling predict the distribution of acoustically voiced and voiceless stop consonants?

(1) In the pre-contrast (or developing contrast) stages of children's speech, it seems to do well, at least for singleton consonants. We have no relevant data about consonant clusters.

(2) In languages with no stop consonant voicing contrast, stops tend to be voiceless in all positions. Clearly, then, we are confronted with a number of languages which maintain articulatorily more difficult stops utterance medially than is necessary. We note that, at the price of this greater articulatory effort, these languages maintain phonetic similarity among their positional allophones. It is as though these languages sacrifice, in part, ease of articulation in favor of limiting acoustic variability in the realization of their segments. A similar case is seen in the speech of those speakers of American English who prevoice their initial /b d g/, which requires an extra articulatory posture but makes such initial stops more like the phonetically voiced medial /b d g/. Thus, careful quantitative modeling identifies a case in which ease of articulation cannot be the only factor in a sound pattern. A value of articulatory modeling is that it identifies those cases that are candidates

for the articulatory ease principle, and those that must require some other principle(s).

(3) In languages with a voicing contrast of some sort, we find few cases of variation. Overall, contrast languages maintain contrasts of articulatorily simpler stops with somewhat more costly stops. In initial position, virtually all languages exploit the easiest category of initial stops, the voiceless unaspirated, as predicted. If our modeling is taken to predict a strong preference for one and only one stop category initially, then of course it fails; but it does predict which of the categories available for contrast will most frequently be chosen by languages. Similarly, in medial position we find no strong preference across languages for any stop consonant category, either the voiced one predicted by modeling, or any other. Some languages favor voiceless unaspirated stops medially, while some favor voiced stops. The model finds more correspondence with natural language in predicting the occurrence of voiceless unaspirated stops in utterance-final position, where neutralization of voicing contrasts is quite common. Limited acoustic data suggests that in some cases at least the details of our prediction -- that final stops should be somewhat, but not greatly, voiced, and therefore acoustically distinct from both completely voiced and completely voiceless stops -- are borne out. Here, however, the lack of acoustic studies of final neutralization of voicing limits our ability to interpret the language survey data with regard to the model's predictions.

Our model provides a forum within which a notion like ease of articulation can be explicitly defined and tested. When the model's predictions and the facts of language disagree, what does that mean? We conclude that, with regard to stop consonant voicing, ease of articulation is probably not the primary determinant of phonetic form, implicitly assuming that our model and our definition of ease of articulation are adequate for our purposes. In this sense, then, our model allows us to find the limits of the influence of ease of articulation on phonetic form, that is, to find those few cases where it seems to play some clear role. We then know which cases remain to be explained, and can hypothesize further principles relevant in such cases, such as communicative efficiency, acoustic invariability, and perceptual requirements.

Footnotes

1. John R. Westbury is currently at the University of North Carolina at Chapel Hill. The research described in this paper was carried out in the Speech Communication Group at M.I.T., under the direction of Prof. Kenneth N. Stevens, with support from post-doctoral fellowships from NIH. We are indebted to Prof. Stevens for his help. Further work and preparation of the manuscript was supported at UNC by the National Institute of Dental Research, and, at UCLA, by the UCLA Academic Senate research program and by a grant from the National Institute of Neurological and Communicative Disorders and Stroke to Peter Ladefoged. We also thank Howard Golub and Gunnar Fant for advice.

2. We can imagine that the ease of articulation principle should consider more than simply the rates of articulatory change between adjacent segments. It may be necessary to specify, for example, how and where transition velocities are to be measured, when those velocities are not constant between sequentially-arranged steady states. Moreover, the principle might consider, and apportion costs on the basis of, the "identities" of state changes. It is plausible, for example, to think that it would be relatively more costly in some metabolic sense to move the rib cage at 100 mm/sec than the tongue tip at twice that speed, simply because of

the great differences in masses of the two structures. Finally, the principle might also consider the "levels" of steady states, between which changes occur. Again, it is plausible to think that sustaining a 25% maximum voluntary contraction (MVC) of a muscle would be relatively easier than one at 50% MVC. Moreover, shifting from 25% MVC to 75% MVC, over a period of 2 seconds, would plausibly be easier than shifting from 50% MVC to 100% MVC over the same period. However, an articulation metric which assigns cost only on the basis of the rates at which state changes occur would consider all states held indefinitely long, and all state changes of the same velocity, to be equally easy. A more general characterization of ease of articulation, suitable for a wide range of speech behaviors, should no doubt consider variables such as these.

3. If we suppose that the vocal folds are fully adducted well before the release of an utterance-initial stop, so that air cannot flow from the lungs to the mouth, then air pressures above and below the glottis will rise asynchronously. Pressure below the glottis will rise first, and pressure above the glottis will begin rising only after the vocal folds have been separated, perhaps after having been "blown apart" once some suitable pressure gradient across the glottis has been reached. Our calculations suggest that 30-40 msec of closure voicing might occur, immediately following the moment when the vocal folds are blown apart and the voicing state is established. Most likely, voicing will also then cease some 30-40 msec prior to release. Thus the closure interval of an initial stop under these conditions would be initially voiceless, subsequently voiced, and then voiceless again, all prior to consonant release. Whether a stop such as that would be considered prevoiced or voiceless, by a phonetician and/or native listener, is an open question. However, we believe that we have occasionally seen stops with this acoustic pattern. Such a scenario would be consistent with the time functions of subglottal and supraglottal pressures during the utterance-initial /b/ illustrated in Figure 2. The utterance-initial /p/ in Figure 3, by comparison, shows subglottal and supraglottal pressures rising synchronously, with the vocal folds no doubt more widely abducted.

References

Baer, T. (1975). Investigation of phonation using excised larynxes. Unpublished doctoral dissertation, MIT.

van den Berg, Jw. (1958). Myoelastic theory of voice production. J. Speech Hear. Res. 1. 227-244.

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

Dinnsen, D. and F. Eckman (1977). Some substantive universals in atomic phonology. Lingua 45. 1-14.

Dinnsen, D. and J. Charles-Luce (1984). Phonological neutralization, phonetic implementation and individual differences. J. Phon. 12. 49-60.

Draper, M., P. Ladefoged, and D. Whitteridge (1959). Respiratory muscles in speech. J. Speech Hear. Res. 2. 16-27.

Eilers, R., D. K. Oller, and C. Benito-Garcia (1984). The acquisition of voicing contrasts in Spanish and English learning infants and children: a longitudinal study. J. Child Lang. 11. 313-336.

Flanagan, J., K. Ishizaka, and K. Shipley (1975). Synthesis of speech from a dynamic model of the vocal cords and vocal tract. Bell Syst. Tech. J. 54. 485-506.

Fujimura, O. and M. Sawashima (1971). Consonant sequences and laryngeal control. Ann. Bul. Res. Inst. Logoped. Phoniat. (Tokyo) 5. 1-6.

Gianinni, A. and U. Cinque (1978). Phonetic status and phonemic function of the final devoiced stops in Polish. Speech Lab. Rep. I, Labotorio di Fonetica Sperimetale (Napoli).

Harms, R. (1978). Some nonrules of English. Linguistic and Literary Studies In Honor of Archibald A. Hill. II: Descriptive Linguistics, ed. by M. A. Jazayery, E. C. Polomé, and W. Winter. 39-51. The Hague: Mouton.

Hirose, H. (1977). Laryngeal adjustments in consonant production. Phonetica 34. 289-294.

Hooper, J. (1976). An introduction to natural generative phonology. New York: Academic Press.

Ishizaka, K., J. French, and J. Flanagan (1975). Direct determination of vocal tract wall impedance. IEEE Trans. Acoust., Speech Signal Process., ASSP-23. 370-373.

Ishizaka, K. and M. Matsudaira (1972). Fluid mechanical considerations of vocal cord vibration. Speech Com. Res. Lab. Monograph 8.

Jakobson, R., G. Fant, and M. Halle (1963). Preliminaries to speech analysis. Cambridge, Mass.: MIT Press.

Keating, P. (1984). Aerodynamic modeling at UCLA. UCLA Working Papers in Phonetics 59. 18-28.

Keating, P. (1983). Physiological effects on stop consonant voicing. J. Acoust. Soc. Am. Suppl. 1, 73. S47. Also in UCLA Working Papers in Phonetics 59, 1984.

Keating, P., W. Linker, and M. Huffman (1983). Patterns of allophone distribution for voiced and voiceless stops. J. Phonetics 11. 277-290.

Ladefoged, P. (1982). A course in phonetics (2nd Edition). New York: Harcourt Brace and Jovanovich.

Ladefoged, P. (1964). Comment on 'Evaluation of methods of estimating subglottal air pressure.' J. Speech Hear. Res. 7. 291-292.

Liljencrants, J. and B. Lindblom (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. Language 48. 839-862.

Lindblom, B. (1983). Economy of speech gestures. The production of speech, ed. by P. F. MacNeilage, 217-246. New York: Springer Verlag.

Lindblom, B. (1978). Phonetic aspects of linguistic explanation. Studia Linguistica 32. 137-153.

Lindqvist, J. (1972). Laryngeal articulation studied on Swedish subjects. Speech Trans. Lab. Q. Prog. Rep. 2-3. 10-27.

Maddieson, I. (1984). Patterns of Sounds. Cambridge: Cambridge University Press.

Muller, E. and W. Brown. (1980). Variations in the supraglottal air pressure waveform and their articulatory interpretation. Speech and language: Advances in basic research and practice, Vol. 4, ed. by N. Lass, 317-389. New York: Academic Press.

Ohala, J. (1983). The origin of sound patterns in vocal tract constraints. The Production of Speech, ed. by P. F. MacNeilage, 189-216. New York: Springer Verlag.

Ohala, J. (1976) A model of speech aerodynamics. Report of the phonology laboratory (Berkeley) 1. 93-107.

Ohala, J. (1974). Phonetic explanations in phonology. Papers from the parasession on natural phonology, ed. by A. Bruck, R. Fox, and M. LaGaly, 251-274. Chigago: Chicago Linguistic Society.

Ohala, J. and C. Riordan (1979). Passive vocal tract enlargement during voiced stops. Speech communicaton papers presented at the 97th meeting of the Acoustical Society of America, ed. by J. Wolf and D. Klatt. New York: Acoustical Society of America.

Preston, M. S., G. Yeni-Komshian, and R. Stark (1967). A study of voicing in initial stops found in the pre-linguistic vocalizations of infants from

different language environments. Haskins Laboratory status report on speech research 10.
Rothenberg, M. (1968). The breath-stream dynamics of simple-released-plosive production. Bibl. Phonetica 6.
Ruhlen, M. (1975). A guide to the languages of the world. Stanford, CA: Stanford University Press.
Schachter, P. (1969). Natural assimilation rules in Akan. Unpublished manuscript.
Schane, S. (1972). Natural rules in phonology. Linguistic change and generative theory, ed. by R. Stockwell and R. Macauley, 199-229. Bloomington, Ind.: Indiana University Press.
Scully, C. (1969). Problems in the interpretation of pressure and air flow data in speech. University of Leeds Phonetics Department Report 2. 53-92.
Smith, B. (1979). A phonetic analysis of consonant devoicing in children's speech. J. Child Lang. 6. 19-28.
Smith, B. (1978). Temporal aspects of English speech production: A developmental perspective. J. Phonetics 6. 37-67.
Smith, N. (1973). The acquisition of phonology. London: Cambridge Univ. Press.
Stampe, D. (1979). A dissertation on natural phonology. Bloomington, Ind.: Indiana Linguistics Club.
Vennemann, T. (1972). Sound change and markedness theory: On the history of the Germanic consonant system. Linguistic change and generative theory, ed. by R. Stockwell and R. Macauley, 230-274. Bloomington, Ind.: Indiana University Press.
Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. J. Acoust. Soc. Am. 73. 1322-1336.
Westbury, J. R. and S. Niimi. (1979). An effect of phonetic environment on voicing control mechanisms during stop consonants. Speech communication papers presented at the 97th meeting of the Acoustical Society of America, ed. by J. Wolf and D. Klatt. New York: Acoustical Society of America.
Zue, V. and M. Laferriere. (1979). Acoustic study of medial /t,d/ in American English. J. Acoust. Soc. Am. 66. 1039-1050.

Lingustic and nonlinguistic effects on the perception of vowel duration

Patricia Keating

0. Introduction

The substantial body of work on categorical perception (reviewed recently by Repp 1983) has produced, as one result, the finding that listeners can divide continua of speech sounds into quite discrete labeling categories, with a fairly sharp boundary between them (Liberman et al. 1957, Liberman et al. 1967, Studdert-Kennedy et al. 1970, Liberman et al. 1972). Though with training listeners can make quite arbitrary categorizations (e.g. Carney et al. 1977), it is often assumed that in the usual case discrete speech perception categories correspond to linguistic phonemes. The result of speech categorization abilities then would be that listeners are able to behave as if largely unaware of allophonic differences, extracting only the intended meaningful units of their language. From this linguistic perspective, we expect listeners' speech perception performance to depend at least in part on their languages' phonemic systems. Listeners' knowledge of how phonetic categories are organized into phonemes should play some role in categorization experiments.

The relevant part of the speech perception literature supports this suggestion. Most obviously, phonemic categories differ in phonetic detail across languages (e.g. Lisker and Abramson 1964 for differences in VOT that were later found to affect perception). Studies on Thai perception (Foreit 1977, Donald 1978) can be interpreted as showing that knowledge of a phonological system is crucial to behavior with adaptation. A more subtle effect of phonemic differences on perception was shown by Keating, Mikoś, and Ganong (1981). They showed that some phonetic category contrasts offer more stable identification performance than other contrasts do.

Given such effects of phonology on speech perception, let us return to the assumption that the categories in speech perception are "phonemes", and that the boundaries between categories are "phoneme boundaries". Exactly what could be meant by "phoneme" here?

Linguistic theories differ in how they view the phoneme. To some it is the unit of surface contrast--a phonetic unit that can change meaning. To others, it is the unit of lexical representation; meaning differences are specified underlyingly in the lexicon, not on the phonetic surface. In this latter view, a surface phonetic contrast that can be derived by rule from some different underlying contrast in the lexicon will not be considered phonemic. An example would be phonetically nasalized vowels that can be derived from sequences of oral vowels plus deleted nasal consonants. Thus, on the surface, there is a contrast between [æ] and [æ̃], as in [kæt] "cat" and [kæ̃t] "can't", but the difference between these words is represented via a different contrast in the lexicon. In this case, then, the "surface" theory posits more phonemes than the "lexical" theory. Conversely, the lexical theory will posit more phonemes in cases where an abstract, absolutely neutralized, phoneme can be justified. An example would be an extra vowel in a skewed vowel harmony system.

However, in very many cases the two kinds of theories lead to positing the same phonemes for a given language. Certainly this has been so for most phoneme

contrasts studied in speech perceptions experiments. When everyone's stimuli consist of [ba da ga pa ta ka], there is no need to worry about which theory of the phoneme to assume: all roads lead to Rome here. At the same time, however, any phonetician with even passing knowledge of phonological theory must feel uncomfortable at the free use of such unexamined terms as "phoneme boundary". Do listeners really hear phonemes, and if so, whose?

In this study the nature of phonemic perception will be investigated by asking how the phonemic status of a phonetic contrast influences categorization. Two cases will be compared: one, a contrast that must be represented in underlying lexical entries, "phonemic" under any theory; the other, a contrast that can be derived by phonological rule and so need not be represented in underlying lexical entries, therefore "phonemic" only for surface-oriented theories. The two cases both involve vowel length contrasts, underlying in Czech and derived in English. In the next section these contrasts are presented in some detail.

I. Vowel length in Czech and English

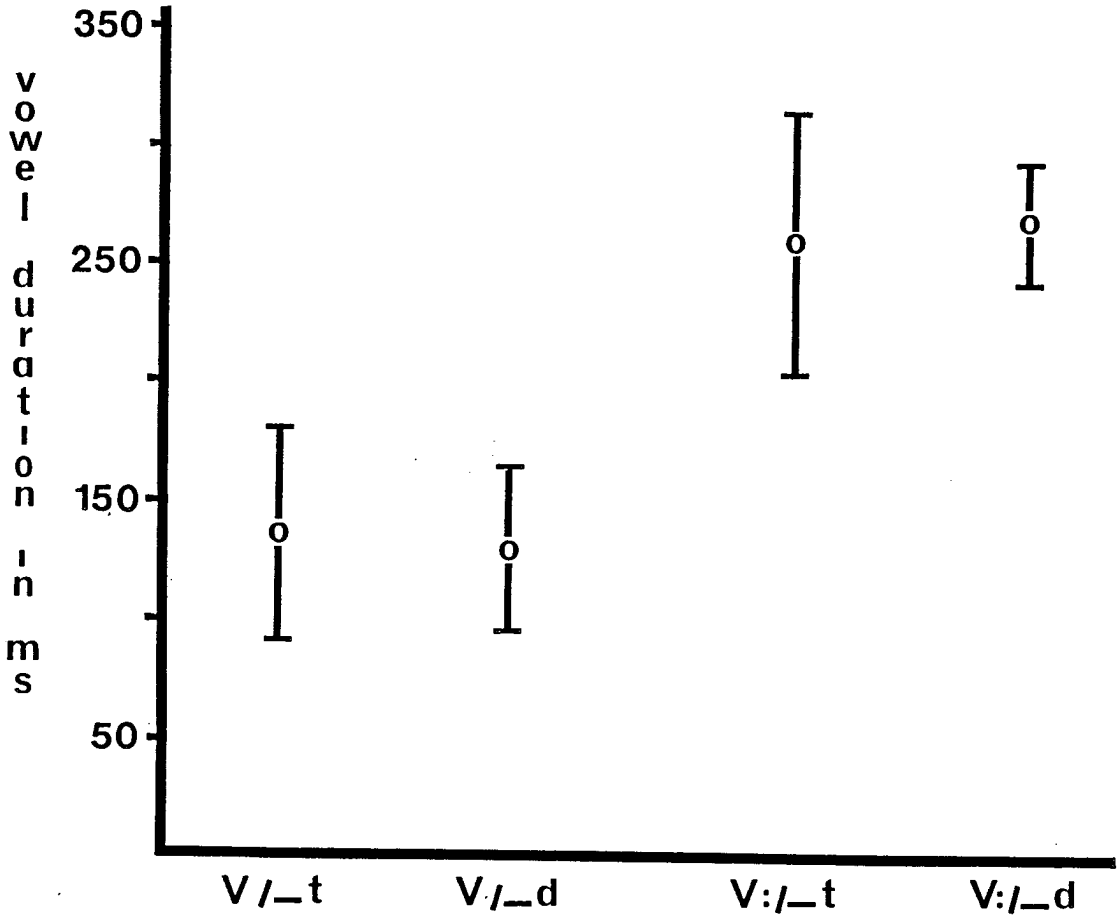
A. Czech

All five of the Czech vowel qualities [i e a o u] occur as both long and short phonemes. Kučera (1961) gives text frequencies of the ten vowels, showing that the short vowels are much more frequent than the long ones. Entirely lacking in his sample is /o:/, which occurs only in loan words. Kučera also indicates that Colloquial Czech has a slightly different vowel system from the standard Literary language. For example, Literary /e:/ corresponds to /e/ or to /i:/ in Colloquial Czech.

Duration is not the only cue to length contrasts in Czech. As might be expected, a certain amount of qualitative difference between long and short vowels is also found. Lehiste (1970) presents acoustic data on these differences, showing that the shorter vowels are more centralized in the vowel formant space. Also as expected, length contrasts are not the only determinants of vowel duration. Phrase prosody also affects measured vowel durations. Kučera (1961) claims that all vowels, both long and short, are lengthened in phrase-final position, thus preserving the phonemic length contrasts. Unlike phrasing, however, lexical stress does not seem to affect vowel duration in Czech as much as in some languages; Lehiste (1970) gives data showing that Czech stressed vowels are no longer than the corresponding unstressed vowels.

Not considered to date in the literature is the effect of consonant voicing on vowel duration. Therefore acoustic measurements of duration were made for long and short vowels before voiced and voiceless consonants. Since Czech has a rule of final obstruent devoicing, vowels before final obstruents were not examined. Instead, vowels /a/ and /a:/ before medial /t/ and /d/ were compared. A total of 32 tokens from two female speakers were recorded and analyzed: 10 short vowels before [t], 6 short vowels before [d], 9 long vowels before [d], and 7 long vowels before [d]. All vowels were in the first (stressed) syllables of real disyllabic words, preceded by either [p] or [ml]. Results are shown in Figure 1. T-tests showed that there was no significant difference in vowel duration between either the short vowel conditions ($t_{14} = .26$) or the long vowel conditions ($t_{14} = -.40$). Thus, perhaps contrary to expectation, Czech vowels are not longer before voiced obstruents, as vowels are in most languages.

Figure 1. Means and standard deviations of durations of Czech vowels before medial /t/ vs. /d/. These differences are not significant.



As far as I know, the perception of vowel length in Czech has not been previously studied. However, perception of vowel length has been studied in other languages. In an experiment of particular relevance to the present study, Bastian and Abramson (1962) had Thai and American listeners label, discriminate, and mimic continua of Thai vowel length differences. The stimuli could be heard as the minimal pair /bàt/ "card" and /bàat/ "monk's bowl". Both groups of listeners showed no discrimination peaks and continuous tracking in mimicry. They differed in labeling performance: the Thais had more discrete labeling categories than the Americans, who were simply responding "short" or "long". The Thai listeners had a fairly sharp boundary halfway along the stimulus continuum.

B. American English

American English vowels show pronounced duration differences of several types. An extensive review can be found in Klatt (1976). For example, the vowels can be divided into two sets often called "tense" and "lax", where both quantity and quality differences are involved. A pair like [i]/[I] differ in that [i] is longer and more peripheral than [I]. The quality difference in English is more extreme, and the quantity difference is less extreme, than in Czech, and generally English is not analyzed as having a phonemic vowel length contrast. In

fact, Stemberger (ms.) presents evidence from English speech errors supporting the view that English vowels are not organized according to length contrasts.

Vowel duration also depends on the segmental context. Vowels are longer before fricatives than before stops, and longer before voiced than before voiceless consonants (thus vowels are longest before voiced fricatives). Reviews of such effects can be found in Lehiste (1970), and most recently, in Eilers et al. (1984). The difference depending on consonant voicing is greatest in monosyllables with only one consonant after the vowel. In this case vowels are about two-thirds as long before voiceless consonants as before voiced. Even in disyllables, vowels before medial consonants show reliable differences in duration, although not as large as in monosyllables. Table 1 summarizes findings in the literature.

Table 1

Published ratios of durations of American English vowels before voiceless obstruents to durations of vowels before voiced obstruents.

Source	Ratio
Monosyllables	
Chen 1970	.61
Zimmerman & Sapon 1958	.64
Peterson & Lehiste 1960	.67
House & Fairbanks	.69
Disyllables	
Sharf 1962	.75
Klatt 1973	.79
Port 1977	.89

Of especial interest has been the relation of vowel duration differences to the American English phonological rule of flapping of /t/ and /d/ before stressless vowels (see Kahn 1976 for discussion of precisely where flapping occurs). Fisher and Hirsh (1976), Fox and Terbeek (1977), and Zue and Laferrière (1979) all found significantly different acoustic vowel durations before flaps derived from /t/ vs. before flaps derived from /d/. In addition, Fox & Terbeek determined that, although 19% of the flaps were phonetically voiceless, the occurrence of such phonetic voicing was uncorrelated with either the underlying stop voicing or with the vowel duration differences. That is, the vowel duration depends on the underlying stop's voicing only. Looking at the phonemic voicing of the underlying stops, Fox and Terbeek found that the vowels before /t/ flaps were 85% the duration of vowels before /d/ flaps. This ratio corresponds to those shown in Table 1 for other disyllables having stops with other places of articulation. However, it should be noted that no such correlation was found by Malécot and Lloyd (1968), who studied Eastern U.S. speakers. Certainly American dialects appear to differ on this point; a number of speakers recorded in the Phonetics Lab at Brown University also failed to show any vowel duration differences, but rather lengthened all vowels before flaps.

Can differences in vowel duration serve as perceptual cues to consonant voicing? Denes (1955), Raphael (1972), and Mermelstein (1978) have found that final consonants can be identified from differences in vowel duration. Port (1977), however, decided from acoustic analysis of vowel duration before medial consonants that vowel duration differences were too slight to serve as perceptual cues. Perception of medial vowel duration as a cue to consonant voicing has not really been studied, with research focused on closure duration instead. Vowel duration before flaps has also been little studied. Lorge (1967) compared words with medial /t/ and /d/, and showed that such words are mis-identified 10% of the time, but she did not indicate when the tokens actually contained flaps. Malecot and Lloyd (1968) did keep track of tokens with flaps and found that they were identified near chance, except when the vowel was /aI/. (In this case quality varies as well as duration: [aI] before /d/-flaps, but [ʌI] before /t/-flaps.) Otherwise, judgments were correlated with vowel duration. However, in both of these studies listeners judged natural tokens, and so variation in duration was not systematically studied.

II. Experiments

A. Design

The goal of the experiments reported here was twofold: first, to extend results on the perception of vowel duration, and second, to explore the phonemic nature of categorization. Therefore the labeling performance of Czech and American listeners was assessed for stimuli varying in vowel duration. One question, then, is simply whether such listeners can reliably categorize such stimuli, particularly whether American listeners can identify flaps with /t/ vs. /d/ on the basis of vowel duration. The other question is which of the two phonemic theories outlined in the opening discussion predicts the pattern of perception. If the "surface" theory is correct, then we predict that perception should reflect surface phonetic contrasts, and that the two languages, with their two types of contrasts, should behave similarly. If the "lexical" theory is correct, then we predict that perception should reflect underlying lexical contrasts, and that the two languages should behave differently.

The experiments involved the usual sort of labeling task: continua of stimuli varying vowel duration in small steps were constructed. Each continuum was made to have endpoint stimuli representing the two vowel duration categories. Typically in a categorization task the stimuli include at least one good exemplar of each category, such that perception of the continuum endpoints is unambiguous. Thus when we find that listeners can label the endpoint stimuli, it is not an interesting result. It is in fact given by the experimental design, through the choice of stimuli, labels, and listeners. Categorization of endpoint stimuli is simply the premise for experimental examination of categorization of acoustically intermediate stimuli: to which category will they be assigned, and how reliably? The usual case in speech perception is that categorization of intermediate stimuli is quite discrete; this can be called the category boundary effect, or categorical labeling. The opposite is continuous labeling, in which categorization changes gradually over several stimuli. Our interest here is in whether vowel duration categorizations will show the category boundary effect. However, in the case at hand of American vowel duration before flaps, mere categorization of endpoint stimuli, by itself, is not guaranteed.

B. Method

1. Subjects

Experiments were run at Brown University in Providence, RI. The Czech listeners were six native speakers, most of them in the Slavics Department, who also knew at least some English. The American listeners were 14 undergraduates from Michigan, Wisconsin, and Illinois. They were unpaid volunteers from among many undergraduates with home addresses in these states to whom solicitations were sent. It was thought that speakers from these states would be most likely to have dialects with the desired vowel duration distinction. Screening to determine participation was post-hoc. These listeners' results were included only if they had reliably categorized the endpoint stimuli in a continuum, i.e., the two most extreme vowel durations. However, a preliminary informal screening of Americans was also carried out by assessing production: when volunteers telephoned, they were asked to pronounce various pairs of spelled words, and only if a vowel duration difference was heard were they scheduled as listeners. There is of course no guarantee that this technique was of any real use, but all of the 14 qualified on at least one listening test; from 10 to 11 qualified on each test.

2. Stimuli

Two Czech and three English vowel duration continua were made for these experiments by editing natural speech tokens produced by a native speaker of each language. The Czech pairs were based on the words lánech and léhat, and the English pairs were based on the words pad and padding. The Czech words were chosen so that both the long and short vowel versions would be real words ("lehat", to lie down; "léhat", to lie; "lanech", ropes (locative, nom. sg. = "lano"); "lánech", expanses (locative, nom. sg. = "lán")). The English words were chosen so that both the /t/ and /d/ versions would be real words, and so that the vowel would be /æ/, which shows little if any acoustic quality changes across durations.

For each continuum, a naturally-produced token with a long vowel was chosen, and digitized at a 10-kHz sampling rate with a 4.5-kHz low-pass filter, with 10-bit quantization, on the PDP 11/34 computer at the Brown University Phonetics Lab. The waveforms of these tokens were viewed, and an editing routine was used to remove selected pitch periods from the vowels. Consonant transitions were not altered in editing; all deletions were from the vowel itself, and the deleted periods were equally distributed over the vowel. Thus, for example, an original token of pad with a duration of 642 msec had 37 steady-state vowel pitch periods, with a total duration of 324 msec, available for editing. The longest stimulus in the continuum was this original token. The shortest token was created from this using pitch periods #1, 6, 11, 16, 21, 26, 31, and 37, together giving a steady-state vowel duration of 65 msec. Because the individual pitch periods varied slightly in duration, the five continua do not have exactly equal step sizes, and because a different speaker is used for each language, the step sizes differ between the two languages. The Czech stimuli have step sizes of roughly 20 msec, or 3-4 pitch periods; the English stimuli, roughly 25 msec, or 2-3 pitch periods. Table 2 gives the number of stimuli, average step size, and range of vowel durations for each. In this table, "vowel duration" measurements are given both in terms of the steady-state durations, and the more usual vocalic nucleus, which includes transitions.

Table 2

Number of stimuli, average step size in msec and in number of pitch periods, and range of vowel durations in msec for each of the five test continua.

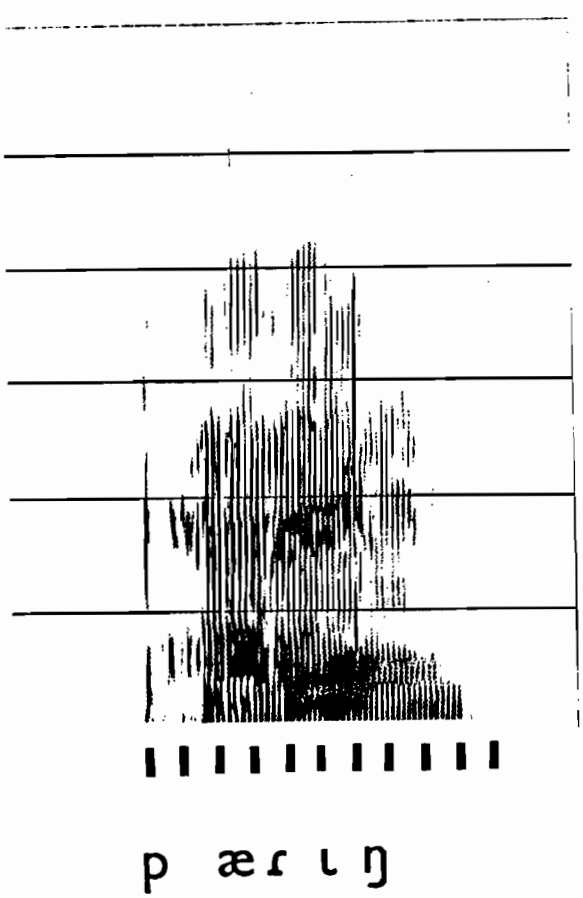
Continuum	# stimuli	average step size	steady-state V duration range	nucleus duration range
pa	11	25	65 - 320 msec	114 - 366 msec
pad	11	25	65 - 320 msec	130 - 382 msec
padding	7	25	45 - 195 msec	106 - 244 msec
lanech	8	20	95 - 235 msec	122 - 260 msec
lehat	8	20	65 - 215 msec	106 - 244 msec

In the case of padding, the longest original token suitable for editing had only 138 msec, in 15 pitch periods, of steady-state vowel available. So that the step size could be comparable across the English continua, extra- long test stimuli were constructed for this continuum only. Selected individual pitch periods were duplicated rather than removed. Thus the longest padding stimulus duplicated pitch periods #3, 4, 6, 8, 11, 12, and 14, giving 195 msec of steady-state vowel duration.

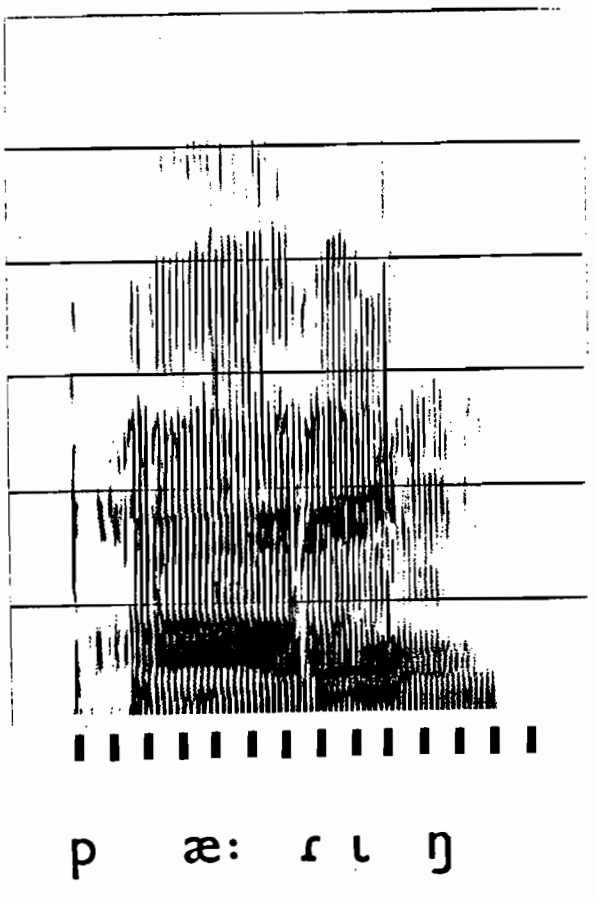
The above describes the treatment of vowel duration. One additional manipulation was performed for the pad continuum. In constructing a continuum from pat to pad, some care must be taken to neutralize consonantal voicing cues. Therefore two pad continua were made, from a single series of vowel duration manipulations. One continuum included the final formant transitions, 62 msec, from the original token. The final consonant was unreleased. This continuum is called the "pad" continuum. The other continuum did not include these transitions. The resulting abrupt vowel offset typically favors a /t/ percept. This continuum is called the "pa" continuum.

In summary, then, five vowel duration continua were made by editing natural tokens. Information about the five stimulus continua is given in Table 2. Figure 2 shows wideband spectrograms of endpoint stimuli from each continuum. For each continuum, a test tape was generated by randomizing 10 repetitions of each stimulus. The tokens were separated by 2 sec, and were grouped into blocks separated by 6 sec.

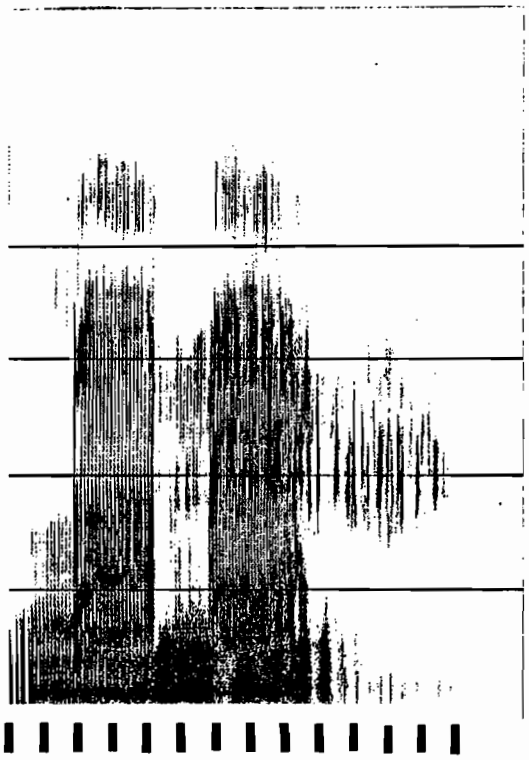
Figure 2. Wideband spectrograms of at least one endpoint stimulus from each continuum. The original Czech words were spoken by a female, and so a 500 Hz bandwidth filter was used to make spectrograms of those tokens. The original English words were spoken by a male, and the usual 300 Hz bandwidth filter was used to make spectrograms of those tokens.



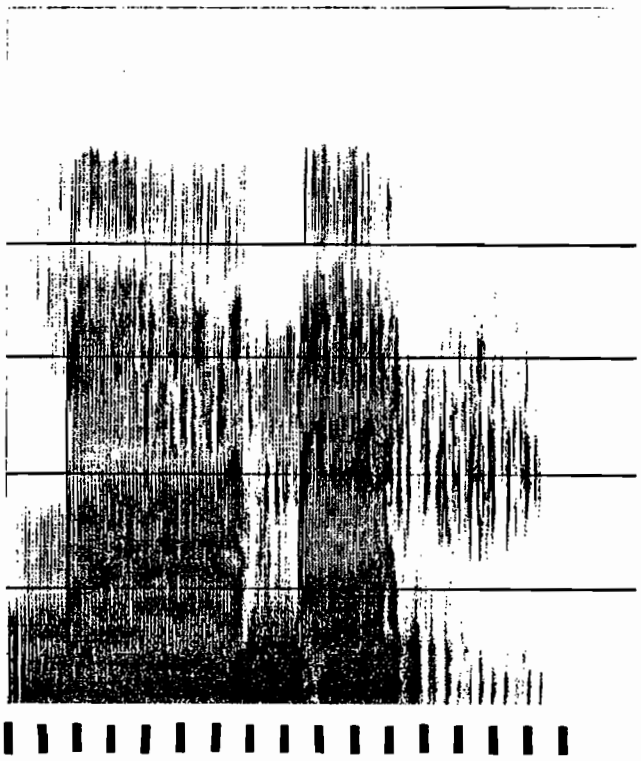
p æɾ ɪ ŋ



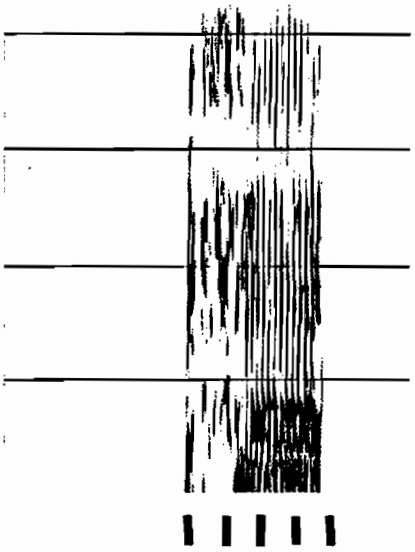
p æː ɾ ɪ ŋ



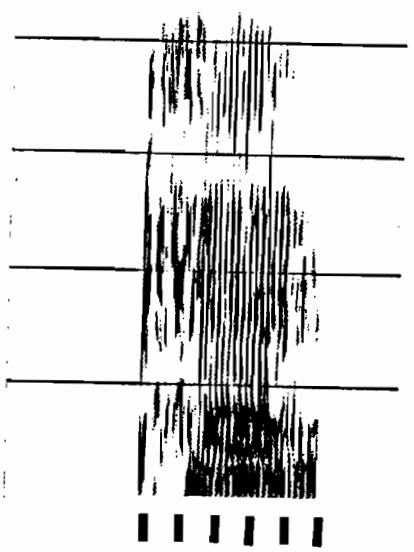
l a n e x



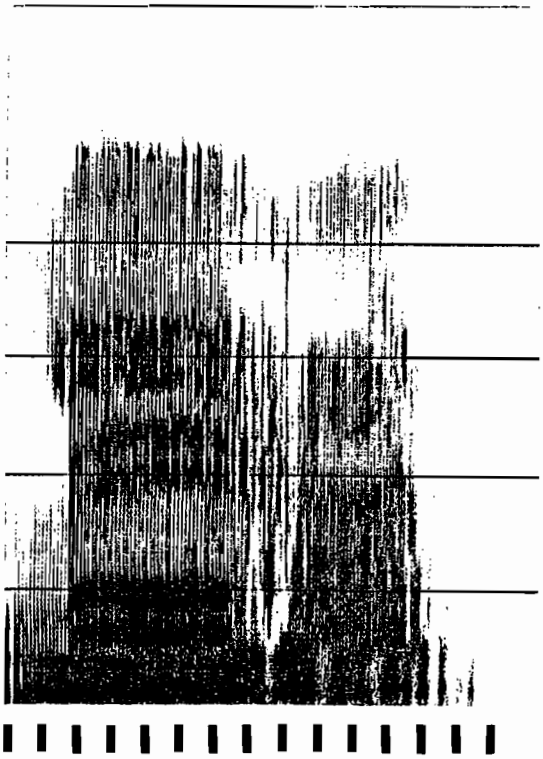
l a: n e x



p æ˥



p æ d˙



l e: ɦ a t

3. Procedure

Each listener heard the set of test tapes for his or her language: two for the Czechs, and three for the Americans. Order of tests was counterbalanced across listeners. Listeners were encouraged to hear the stimuli as real words, and they provided real-word answers in a forced-choice format by putting a check mark in the appropriate column on answer sheets. Note that the Americans were choosing between answers that differed orthographically (and phonologically) in the voicing of the second consonant, though they were hearing phonetic differences of vowel duration. The phonetics of the stimuli were not discussed with the listeners, and none claimed to have any trouble matching stimuli with answer categories.

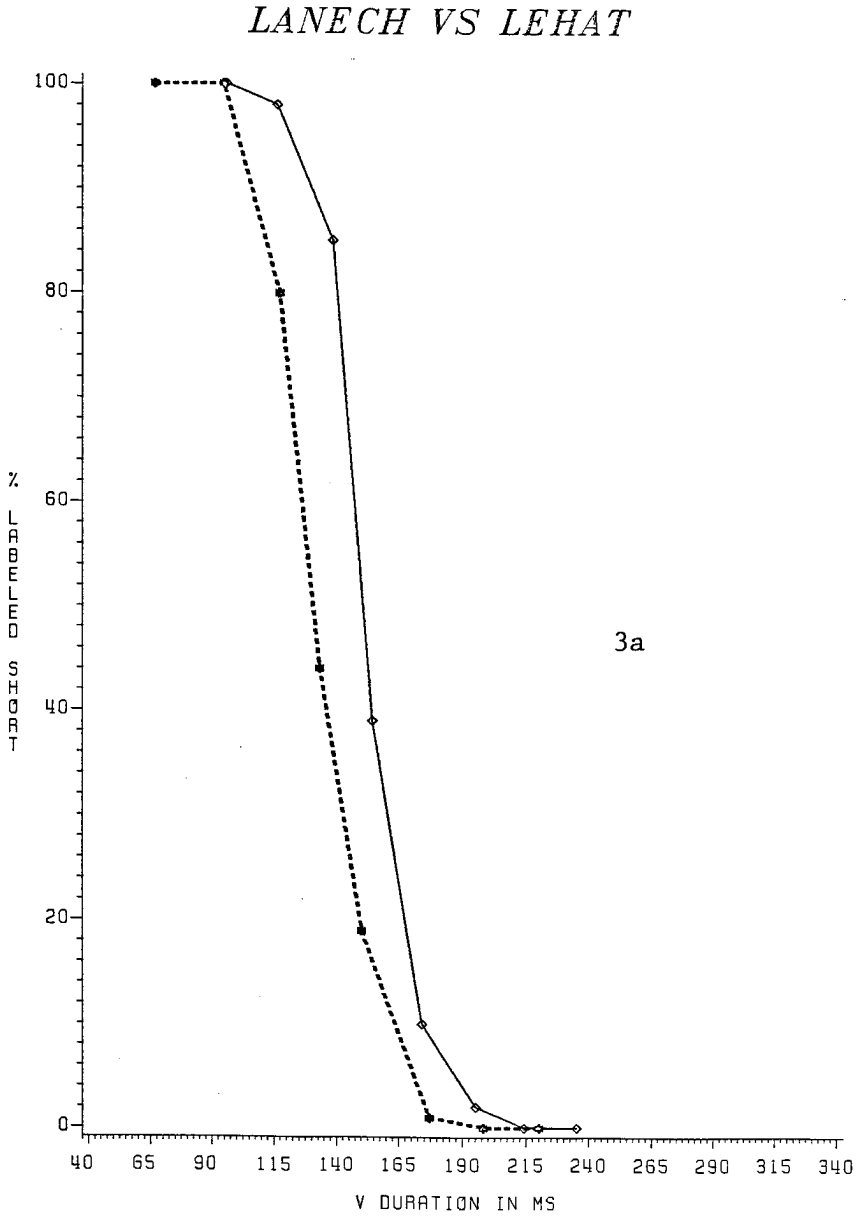
In addition, seven of the fourteen Americans listened to the Czech lánech test tape after completing the three English tapes, thus replicating Bastian & Abramson's (1962) Thai experiments. They were told to hear the stimuli as "short" and "long", and listened to a few practice items before beginning the test. All seven felt reasonably comfortable performing in this part of the experiment.

C. Results and Analysis

The categorization of each stimulus was tabulated for each listener. These data were inspected to determine which listeners to include in statistical analyses. The criterion for inclusion was at least 80% of the shortest stimulus tokens assigned to one category, and at least 80% of the longest stimulus tokens assigned to the other category. That is, the endpoint stimuli had to be consistently assigned to the two categories. With this criterion, all of the Czech data was included, but one to four Americans were eliminated per test. Ten out of fourteen Americans were included for the padding and pa (no transitions) continua, eleven out of fourteen were included for the pad (unreleased transitions) continuum, and six out of seven Americans were included for the Czech lánech continuum. Thus, to answer two of our motivating questions, first, most but not all of the American listeners could respond to vowel duration differences to label consonant voicing differences. While some listeners obviously had difficulty with this task, they had no more difficulty with the flapping cases than with the monosyllable cases. Second, Czech listeners could easily label these continua as long and short vowel categories. The mean identification functions for the five tests are shown in Figure 3.

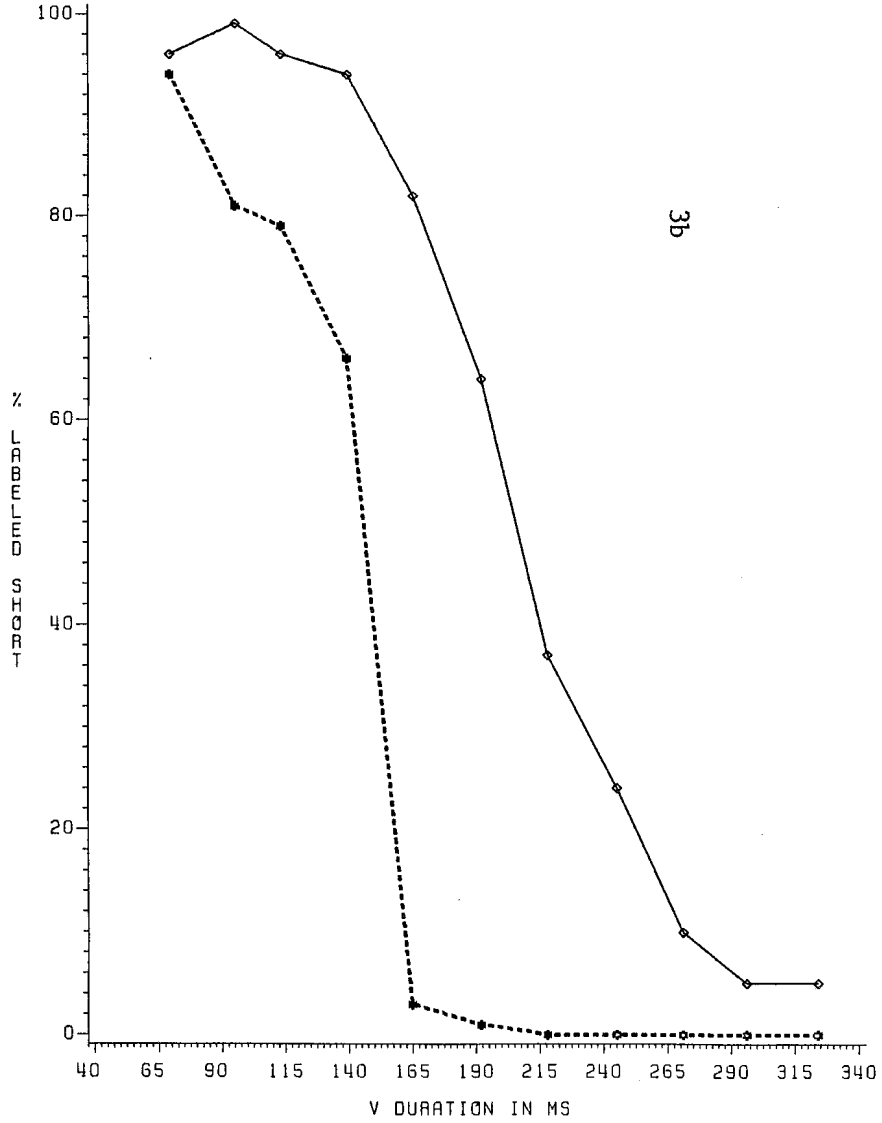
For each of the fifty sets of labeling responses that met the inclusion criterion, three measures were computed. All are concerned with the shape of the labeling function between the endpoint stimuli. Two measures were computed using Probit analysis (Finney 1971), which fits a straight line to z-score transforms of percentage responses to each stimulus. First, the vowel duration value in msec where the fitted line has a labeling value of 50%--the category boundary or crossover point--was determined. Then the slope of the fitted line was computed. The third measure was not based on Probit analysis. This was the "boundary width", the span of the stimulus continuum for which stimuli were not labeled as belonging to one category for at least 80% of their tokens. The slope and the boundary width are both measures of how discrete the two categories are -- how sharp the boundary between them is. A sharply falling identification function indicates a strong category boundary effect. The boundary value indicates how the continuum is divided into categories.

Figure 3. Mean labeling performance for each of five tests. Figure 3a shows the Czech functions: lanech as diamond points connected by a solid line, and lehat as stars connected by a dotted line; Figure 3b shows the American final consonant functions: pa (no final transitions) as diamond points connected by a solid line, and pad (final transitions) as stars connected by a dotted line; Figure 3c shows the American flap function, padding.



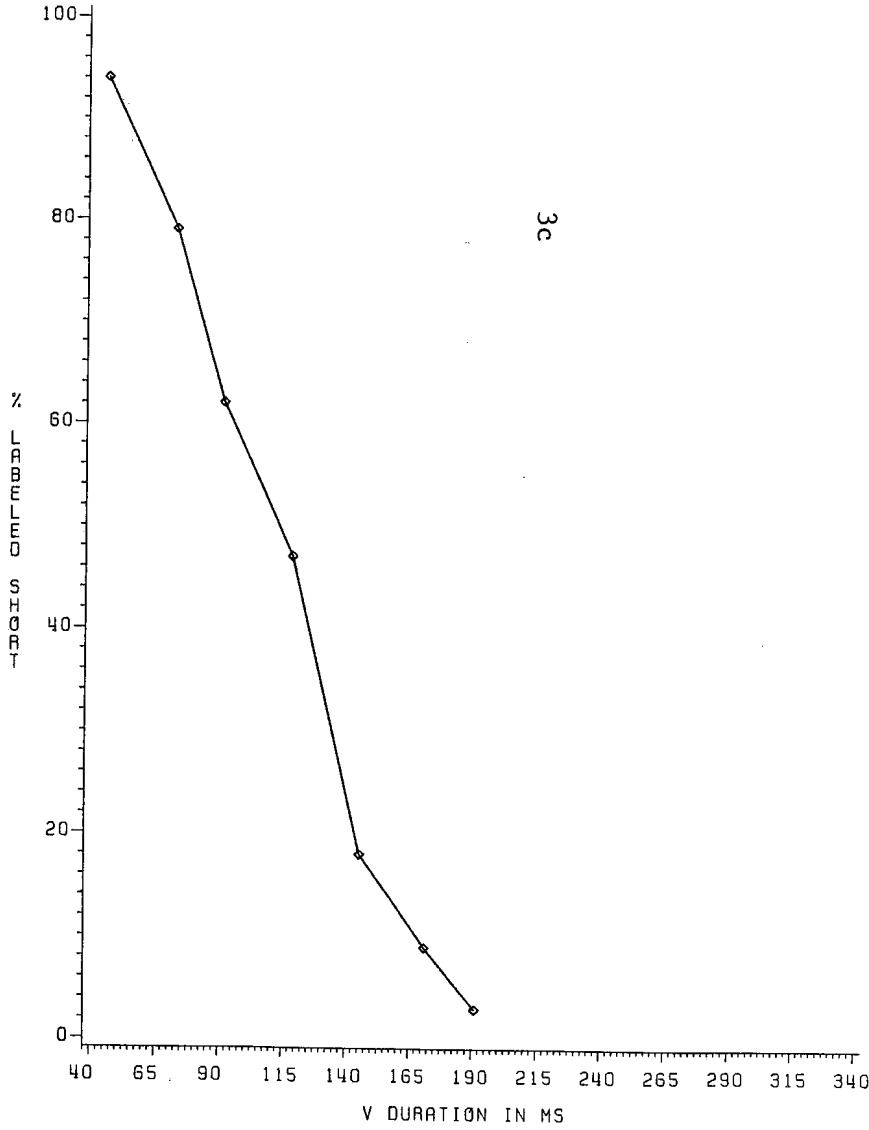
3a

PA VS PAD



3b

PADDING



3c

Recall that the continua from the two languages differ in their step sizes because of the different speakers used. The Czech continua have smaller step sizes in time, but larger step sizes in number of pitch periods. When people hear duration, do they hear absolute time, or are they more sensitive to how many pitch periods are contained in that time? To allow for both possibilities, the boundary width analysis was done two ways: in msec time, and in number of pitch periods.

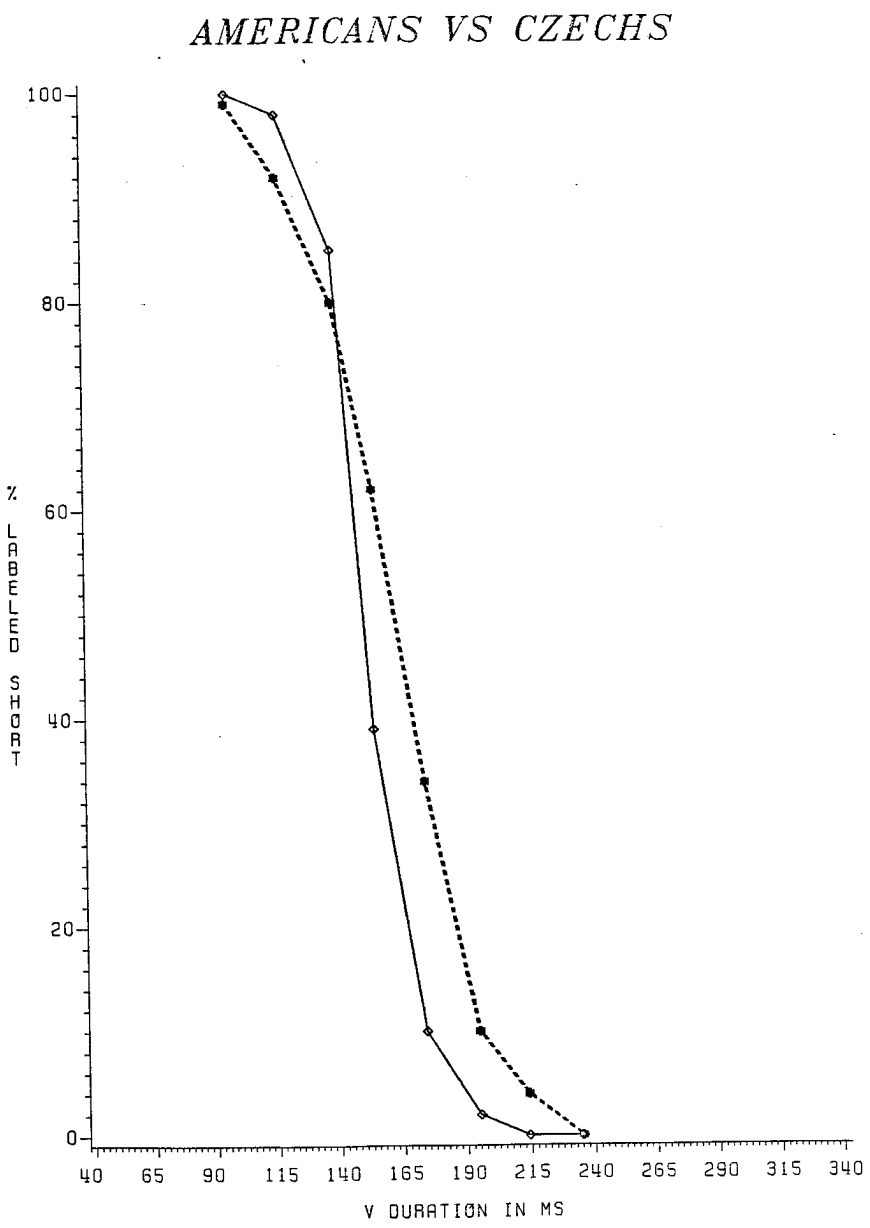
These measures were compared across test continua using two-tailed t-tests. In the case of boundary values, only comparisons of same-vowel continua are relevant, i.e. American pad and pa, with and without final transitions. The mean boundary values were 144 and 209 msec respectively, a significant difference (t_{19} = 5.45, $p < .001$). As expected, then, stimuli without final /d/ transitions require much longer vowel durations for a final /d/ to be heard. In the case of boundary widths and slopes, the two Czech tests were compared, with no significant differences, and the three American tests were compared pairwise, again with no significant differences. For the boundary width comparisons no difference was found with any of the methods of calculating boundary width. Thus listeners of each language performed at a consistent level across tests in their ability to assign the stimuli to discrete categories.

The real question, however, is how the Czech listeners compare with the Americans in discreteness of categories. Therefore all the Czech boundary widths and slopes were compared against all the American ones. All but one comparison indicated statistically significant differences. With boundary widths expressed in msec, the difference between the Czechs (mean value of 40.2 ms) and the Americans (mean value of 83.7 ms) is significant (t_{41} = 4.45, $p < .001$). With boundary widths expressed in number of pitch periods, the difference between the Czechs (mean value of 7.75) and the Americans (mean value of 9.29) is not significant (t_{41} = -1.453). But with just the padding condition compared against the Czech data, a significant difference is seen. With boundary widths expressed in msec, the difference between the Czechs (mean value of 40.2 ms) and the Americans (mean value of 93.0 ms) is significant (t_{20} = 7.11, $p < .001$). With boundary widths expressed in number of pitch periods, the difference between the Czechs (mean value of 7.75) and the Americans (mean value of 8.2) is again significant (t_{20} = -2.70, $p < .02$). The slopes were calculated in terms of msec vowel duration. The average Czech slope is almost three times the average American slope, a much greater difference than could be due to differences in step size alone (-.08 vs. -.03). This difference is significant (t_{41} = 8.51, $p < .001$). Thus with all but one of the measures used to assess sharpness of labeling function, the Czech categorize their stimuli more discretely than the Americans do.

Recall that half of the Americans also listened to the Czech lánech continuum. A comparison of the Czech and American results for this continuum is shown as mean functions in Figure 4; statistical comparisons were also made. The difference in boundary values between the Czechs (mean value of 152 msec) and the Americans (mean value of 153 msec) is not significant (t_{10} = -.07). That is, both groups divided this continuum into categories at the same place. However, the measures of discreteness show the same kind of differences as when the two languages are compared. The difference in slopes between the Czechs (mean value of -.08) and the Americans (mean value of -.04) is signigicant (t_{10} = 2.66, $p < .05$). With boundary widths expressed in msec duration, the difference between the Czechs (mean value of 40 ms) and the Americans (mean value of 62 ms) is significant (t_{10} = -2.29, $p < .05$). With boundary widths expressed in number of

pitch periods, the difference between the Czechs (mean value of 7.7) and the Americans (mean value of 11.7) is significant ($t_{10} = -2.25$, $p < .05$). This effect is apparent in Figure 4 as well.

Figure 4. Mean labeling performance of Czech vs. American listeners for the Czech lanech continuum. The Czech function is shown as diamond points connected by a solid line, and the American function is shown as stars connected by a dotted line.



To summarize the results of these experiments: First, some Americans can reliably use vowel duration as a cue to the voicing of a final unreleased consonant, and as a cue to the source of flaps. Second, when they do this, their labeling functions are less steep -- the categories are less discrete -- than the functions of Czechs judging lexical vowel length differences. Third, when these Americans listen to the same stimuli as the Czechs, their functions are again less steep, but surprisingly, their category boundaries are the same.

III. Discussion

The experiments reported here have shown a difference in performance between Czech and American listeners in identification of stimuli varying in vowel duration. The results of these experiments can be viewed as showing both linguistic and nonlinguistic effects on speech perception. Where the two groups of listeners perform differently, we can see linguistic effects. Where they perform similarly, with no basis in similarities between their languages, then we can see nonlinguistic effects.

Comparison of the Czech and American results in categorization showed the Czechs to have more discrete categories than the Americans. Is there a linguistic difference between the two languages that could account for this difference in performance? Both English and Czech have long and short vowels in their phonetics, but the phonological status of those length contrasts is different in the two languages. The Czech contrast is present in lexical entries, while the English contrast is derived by phonological rule. A possible conclusion, then, is that precisely that difference in phonemic status causes the difference in categorization performance. The difference observed in discreteness of vowel length categories is a linguistic effect on perception. We asked which type of phoneme, surface or lexical, is represented in categorization performance. The conclusion here would be that the categories of categorization may ideally be lexical, not derived surface, contrasts.

The intuition behind this interpretation is an appealing one. By saying that the vowel duration contrast in English is derived, we encode the idea of an indirect connection between what a listener hears and what he can access in lexical representations. The English listener must in some sense translate phonetic vowel duration differences into phonemic stop voicing differences. The results of this experiment indicate that such a process introduces some uncertainty into categorization performance. The Czech categorizations, where no such translation is required, are more consistent than the American ones.

An alternative interpretation might be that the functional load of the vowel duration contrast differs in the two languages, and that the experimental result is due to this difference. It does seem plausible that derived contrasts would usually have a lower functional load than underlying contrasts. It must also be remembered that distinctive vowel length before flaps is dialectally limited in American English. Thus perhaps American listeners are simply less experienced with their experimental contrast than are the Czech listeners. However, vowel length contrasts before final unreleased stops, with the final voiced stops often devoiced, should have a high load in most dialects of English, with many types and many tokens. Yet the continua representing these contrasts showed the same degree of categorization as the more limited flapping contrast. Furthermore, if one claims that functional load is a factor in this result, one would also predict that other low-load lexical contrasts, such as English /s/ vs. /θ/), would show the same sort of categorization. Such a prediction seems unlikely.

The other principal result of the experiment was that the Americans divide the Czech stimuli into categories at the same point along the vowel duration continuum as do the Czech listeners. The similar location of the category boundaries for this continuum suggests a nonlinguistic basis for the categorization of vowel duration contrasts. Presumably the auditory system would constrain the perception of duration contrasts in some way that forces just this categorization. Possibly most, if not all, of the major phonetic categories used by languages have some psychophysical basis in the auditory system (Kuhl & Miller 1978, Pisoni 1977, Blumstein & Stevens 1979). If this is the case, we might expect infants and even animals to perform similarly on these stimuli. Eilers et al. (1984) have shown that American infants can discriminate vowel duration differences. Kuhl (1981) discusses how mammalian auditory constraints could in general select for particular categories. (For reviews of the animal and infant perception literature, see Kuhl 1978, 1979).

However, although the location of the boundary was the same for the Czech and American listeners of the Czech continuum, the categories of the Czech listeners were more discrete. The same sort of result was obtained by Bastian & Abramson in their comparison of Thai and American listeners to a Thai vowel length contrast. Although Bastian & Abramson did not assess function shape quantitatively, it appears from their Figure 3 and their discussion that the Americans provided less discrete categorizations than did the Thais. These results suggest that linguistic experience plays a role in sharpening up an inherent ability to distinguish categories along a phonetic dimension. On this point, we can consider comparisons of humans and animals on other phonetic contrasts. Chinchillas have boundaries for consonant voicing and place of articulation very much like humans, despite their lack of relevant linguistic experience. However, the chinchillas' categorization is more continuous than the humans' (Kuhl and Miller 1978, Kuhl and Padden 1983). Thus, the location of category boundaries could have a nonlinguistic basis, but linguistic experience would result in sharpening of such boundaries.

In conclusion, a comparison of how Czech and American listeners label stimuli differing in vowel duration shows both linguistic and nonlinguistic effects on perception. The linguistic effect is that discreteness of labeling categories differs across the two languages. This difference suggests that the "phonemes" of the "phoneme boundary effect" may be underlying lexical phonemes, rather than derived surface phonemes. The nonlinguistic effect is that the location of the category boundary for both sets of listeners judging Czech stimuli is the same. This similarity suggests that vowel duration contrasts may have an inherent, auditory basis.

Acknowledgment

The experimental work was carried out at the Brown University Phonetics Lab, and was made possible through the programming efforts of John Mertus. I would like to thank him, the volunteer subjects in the experiments, Profs. Henry Kucera and W. Nelson Francis for discussion of features of Czech and American use of vowel length, Elan Dresher for the original idea and encouragement, W. Francis Ganong for help with the experiments, and finally Bruce Hayes, Peter Ladefoged, and Aditi Lahiri for comments on the manuscript.

REFERENCES

BASTIAN, J. and ABRAMSON, A. S. (1962). Identification and discrimination of Phonemic Vowel Duration. J. Acoust. Soc. Am. 34 (5), 743-4 (A).

BLUMSTEIN, S. E., and STEVENS, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of step consonants. J. Acoust. Soc. Am., 66, 1001-1017.

CARNEY, A. E., WIDIN, G. P., and VIEMEISTER, N. F. (1977). Noncategorical perception of stops differing in VOT. J. Acoust. Soc. Am., 51, 483-502.

CHEN, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. Phonetica, 22, 129-159.

DENES, P. (1955). Effect of duration on the perception of voicing. J. Acoust. Soc. Am. 27, 761-764.

DONALD, S. L. (1978). The Perception of Voicing Contrasts in Thai and English. Ph.D. Diss, The University of Connecticut.

EILERS, R. E., BULL, D. H., OLLER, D. K., and LEWIS, D. C. (1984). The discrimination of vowel duration by infants. J. Acoust. Soc. Am., 75, 1213-1218.

FINNEY, D. J. (1971). Probit Analysis. Cambridge U. P., Cambridge.

FISHER, W. M. and HIRSH, I. J. (1976). Proceedings of the 12th Meeting, Chicago Linguistic Society.

FOREIT, K. G. (1977). Linguistic relativism and selective adaptation for speech: a comparative study of English and Thai. Percept. and Psychophysics, 21(4), 347-351.

FOX, R. and TERBECK, D. (1977). Dental flaps, vowel duration and rule ordering in American English. J. Phonetics, 5, 27-34.

HOUSE, A. and FAIRBANKS, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. J. Acoust. Soc. Am., 25, 105-113.

KAHN, D. (1976). Syllable-based generalizations in English phonology. Indiana U. Linguistics Club.

KEATING, P., MIKOŚ, M. J., and GANONG, W. F. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. J. Acoust. Soc. Am., 70(5), 1261-1271.

KLATT, D. (1973). Interaction between two factors that influence vowel duration. J. Acoust. Soc. Am., 1102-04.

KLATT, D. H. (1976). Linguistic uses of segmental duration in English: acoustic and perceptual evidence. J. Acoust. Soc. Am., 59, 1208-1221.

KUČERA, H. (1961). The Phonology of Czech. Mouton (The Hague).

KUHL, P. K. (1978). Predispositions for the perception of speech-sound categories: A species- specific phenomenon? in Communicative and cognitive abilities--Early behavioral assessment, ed. by Minifie, F. D. and Lloyd, L. L., University Park Pr., 229-255.

KUHL, P. K. (1979). The perception of speech in early infancy. in Speech and language: Advances in basic research and practice, I, ed. by Lass, N. J., Academic Pr., 1-47.

KUHL, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. J. Acoust. Soc. Am., 70(2), 340-355.

KUHL, P. K. and MILLER, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. J. Acoust. Soc. Am., 63, 905-917.

KUHL, P. K. and PADDEN, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. J. Acoust. Soc. Am., 73(3), 1003-1010.

LEHISTE, I. (1970). Suprasegmentals (MIT, Cambridge, MA).

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D., and STUDDERT-KENNEDY, M. (1967). Perception of the speech code. Psychol. Rev., 74, 431-36.

LIBERMAN, A. M., HARRIS, K. S., HOFFMAN, H. F., and GRIFFITH, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. J. Exp. Psychol., 61, 379-388.

LIBERMAN, A. M., MATTINGLY, I. G., and TURVEY, M. T. (1972). Language codes and memory codes. Coding processes in human memory, ed. by A. W. Melton and E. Martin. Winston, Washington DC, 307-334.

LISKER, L., and ABRAMSON, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. Word, 20, 384-422.

LORGE, B. (1967). A study of the relationship between production and perception of initial and intervocalic /t/ and /d/ in individual English speaking adults. Status Report SR-9, Haskins Laboratories, pp. 3.11-3.18.

MALECOT, A. and LLOYD, P. (1974). The /t/:/d/ distinction in American alveolar flaps. Lingua, 19, 264-272.

MERMELSTEIN, P. (1978). On the relationship between vowel and consonant identification when cued by the same acoustic information. Percept. Psychophys., 23, 331-36.

OHALA, J. (1974). Phonetic Explanation in Phonology. CLS Parasession on Natural Phonology.

PISONI, D. B. (1977). Identification and discrimination of the relative onset of two component tones: Implications for the perception of voicing in stops. J. Acoust. Soc. Am., 66, 30-45.

PORT, R. (1977). The influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words. Ph.D. diss., U. Conn., published by the Indiana U. Linguistics Club.

RAPHAEL, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in English. J. Acoust. Soc. Am., 51, 1296-1303.

REPP, B. H. (1983). Categorical perception: Issues, Methods, Findings. in Speech and language: Advances in basic research and practice, Vol. 9, ed. by Lass, N. J., Academic Press.

SHARF, D. (1962). Duration of post-stress inter-vocalic stops and preceding vowels. Language and Speech, 5, 26-31.

STEMBERGER, J. P. (1983). Length as a Suprasegmental: Evidence from Speech Errors. Unpublished ms., Carnegie-Mellon University.

STUDDERT-KENNEDY, M., LIBERMAN, A. M., HARRIS, K. S., and COOPER, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review. Psychological Review, 77, 234-249.

ZIMMERMAN, S. and SAPON, S. (1958). Note on vowel duration seen cross-linguistically. J. Acoust. Soc. Am., 30, 152-3 (L).

ZUE, V. W. and LAFERRIERE, M. (1979). Acoustic study of medial /t,d/ in American English. J. Acoust. Soc. Am., 66(4), 1039-1050.

Hausa vowels and diphthongs.

Mona Lindau-Webb

Phonetics Laboratory, Department of Linguistics, UCLA, and
Department of Linguistics and Phonetics, University of Lund, Sweden.

In this paper acoustic data on vowels and diphthongs in Hausa are presented. Two topics are discussed in relation to these data, namely the relation between long and short vowels, and the nature of the diphthongs, as well as what kind of model can best describe the diphthongs.

Hausa belongs in the Chadic branch of the Afroasiatic language family. Chadic languages in general are characterized by few contrasting vowel phonemes with an abundance of environmentally conditioned allophones. Hausa has five contrasting long vowels: ii, ee, aa, oo, uu, and five short vowels: i, e, a, o, u. In addition there are two diphthongs, /ai/, and /au/. The occurrence of short /e/ and short /o/ in non-final position is marginal, and restricted to loanwords, e.g. from English. These vowels are not included in the study. The phonemic contrast between short /i/ and short /u/ in non-final position is questionable. Parsons (1970) and Schuh (1971) suggest that in non-final position the only contrasting short vowels are a low /a/ and a high vowel that varies between [i] and [u], depending on environment. Following Salim (1977) and Newman (1979) I will consider short /i/ and short /u/ to be phonemes. In addition, Hausa is usually described as having two diphthongs, /ai/ and /au/.

The Hausa syllable structure is fairly simple. A syllable can be heavy: CVV, or CVC; or light: CV. Vowel length is contrastive, but this contrast is restricted to open syllables. In closed syllables only short vowels can occur. Long geminated consonants occur phonetically across contiguous syllables, when the syllable-final consonant in one syllable is identical to the syllable-initial consonant in the following syllable in sequences of the type CVC-CV.

Procedure.

A set of disyllabic words was selected to illustrate vowels and diphthongs in the first syllable. The data set was not designed to cover vowels in all possible positions. The environments for the vowels were mainly restricted to initial labial and alveolar consonant(s) and following medial alveolar consonant(s), which were in turn followed by /a/. The mid vowels occurred only after alveolar consonants. The selected items thus consisted of real disyllabic words of the type:

$$\{^{\text{labial C}}_{\text{alveolar C}}\}\ \text{V(V) alveolar C(C) a(a)}$$

These words were put in a frame:

/ban cee CVCV ba/ "I did not say..." or
/ban cee yaa CVCV ba/ "I did not say he...(verb)"

In addition, a limited set of vowels after initial palatal and velar consonants were included. Most of the vowels and diphthongs were on high tones.

Ten speakers of Kano Hausa recorded these utterances in Kano. Measurements were made from wideband spectrograms of the durations and formant frequencies of the vowels. All measurements were made twice, six months apart. Vowel and diphthong durations were measured in milliseconds as the duration of the first formant. The durations of the steady state and transitional parts of the diphthong /au/ were measured from the second formant. Durations of the intervocalic medial plosives were also measured. This duration was taken to be the closure plus the release. Paired t-tests were used to determine the statistical significance of durational differences. Formant frequencies were measured from steady states of the vowels. If a vowel contained no steady state, its frequencies were taken from the middle of the vowel. The formant frequencies were measured in Hertz, then converted to mel to better correspond to perceptual distances. The formant frequency values in mel were then plotted on an acoustic chart with F2 against F1. Ellipses with centers on the mean F1 and F2 for each vowel and with radii of two standard deviations were drawn along axes that were oriented along the principal components of each vowel distribution. Such an ellipse encompasses approximately 95 % of the variation in each cluster.

Results and discussion.

Length.

The durations of high and low vowels for the ten speakers were studied in the environment before alveolar plosives. The data consist of long /ii/ and /aa/, and short /i/ and /a/ in both open and closed syllables before voiceless alveolar plosive /t/. The low vowel /a/ was measured in the environment before voiced alveolar plosive /d/ as well. The data came from words with the following syllable structures: CV, CV-CV, CVC-CV. The mean durations of the vowels and the alveolar consonants of the ten speakers have been plotted in figure 1.

Average durations in milliseconds of long and short /ii/ and /i/ before /t/, and long and short /aa/ and /a/ before /t/ and /d/ for ten speakers are listed in Table 1.

msec.	/ii-t/	/i-t/	/it-t/	/aa-t/	/a-t/	/at-t/	/aa-d/	/a-d/	/ad-d/
$\bar{x}$	106	67	46	118	71	61	127	70	52
SD	18	16	15	15	8	6	16	11	12
	n=10								

Table 1.
Means and standard deviations of the durations of /ii/ and /i/ before voicless alveolar plosive, and /aa/ and /a/ before voiceless and voiced alveolar plosive for ten speakers. The short vowels in open and closed syllables are listed separately.

Hausa has three significantly different phonetic vowel lengths: long vowels; short vowels in open syllables; and short vowels in closed syllables. In open syllables the durational differences between long and short vowels are quite large. Roughly speaking, the long vowels are about 40-45% longer than the short

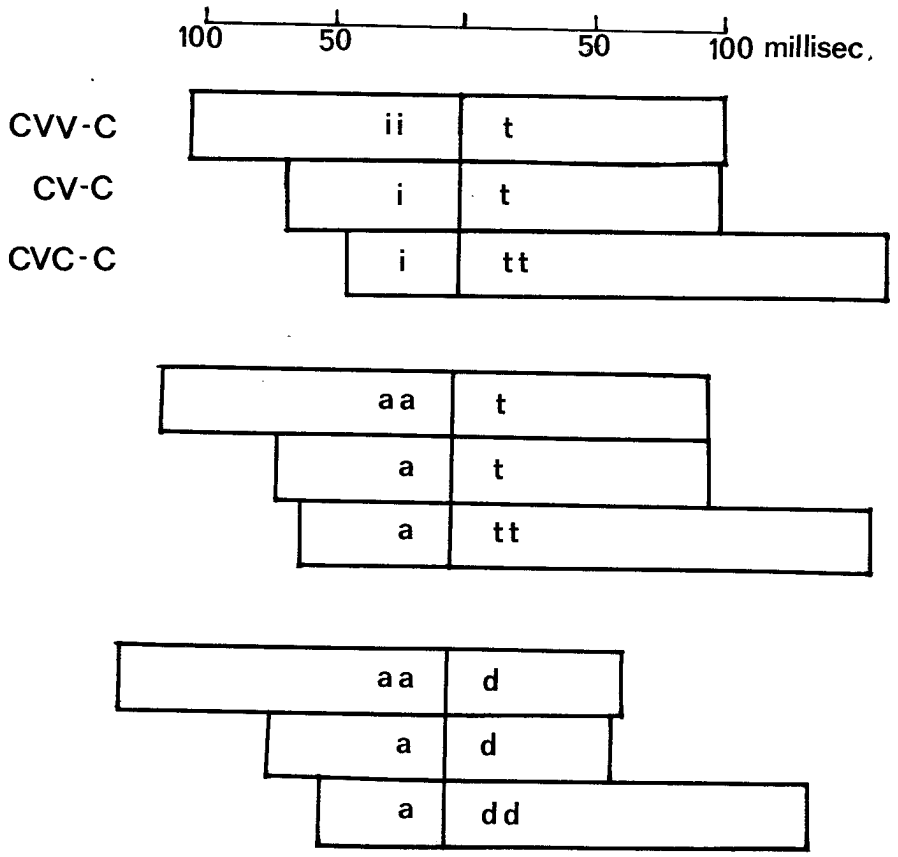


Figure 1. Bargraphs of the mean medial vowel and consonant durations of ten speakers in the three syllable types CVV-C, CV-C, and CVC-C.

vowels. The differences in duration between short vowels in open syllables and short vowels in closed syllables are less than those between long and short vowels in open syllable, but they are still significant. ($p < 0.001$). The open syllable short vowels are about 15-20% longer than the ones in closed syllables.

In many languages there exists a compensatory relationship between vowel length and voicing in the following consonant. Vowels tend to be shorter before voiceless than before voiced consonants, and postvocalic voiceless consonants are usually longer than voiced consonants. In Hausa, however, the situation is more complex. As in most languages the postvocalic voiceless plosives are significantly ($p<0.001$) longer than their voiced counterparts. One might expect that at least the vowels in closed syllables should exhibit vowel shortening before voiceless consonants. But instead only the long vowels in open syllable follow the expected pattern and are significantly ($p<0.01$) shorter before voiceless than before voiced plosives. The durations of the short vowels do not differ significantly before voiced and voiceless consonants. In fact, if anything there is an unexpected weak tendency for the opposite relationship in that the short vowels in closed syllable tend to be longer before voiceless than before voiced consonants ($p<0.05$). It has been proposed that the compensatory relationship between vowels and voicing in postvocalic consonants is universal, and therefore not part of the grammatical rules of individual languages (Chomsky and Halle 1968). But there is now a considerable body of evidence that this relationship is not found in all languages, and that it can be highly language-specific (e.g. Port et al. 1980). The results from Hausa support the view that it is necessary to have language-specific rules for temporal

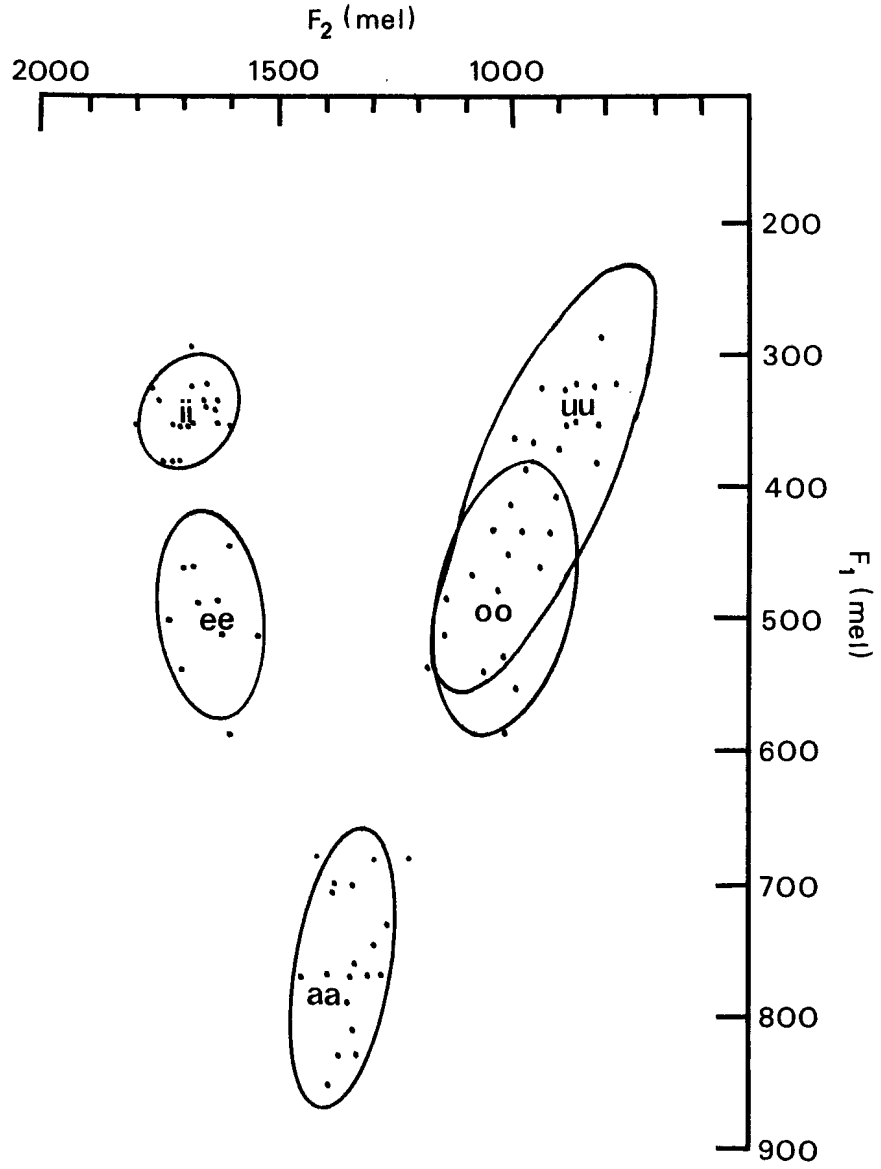


Figure 2. The formant space on a mel scale for the long vowels /ii/, /ee/, /aa/, /oo/, and /uu/ for ten speakers.

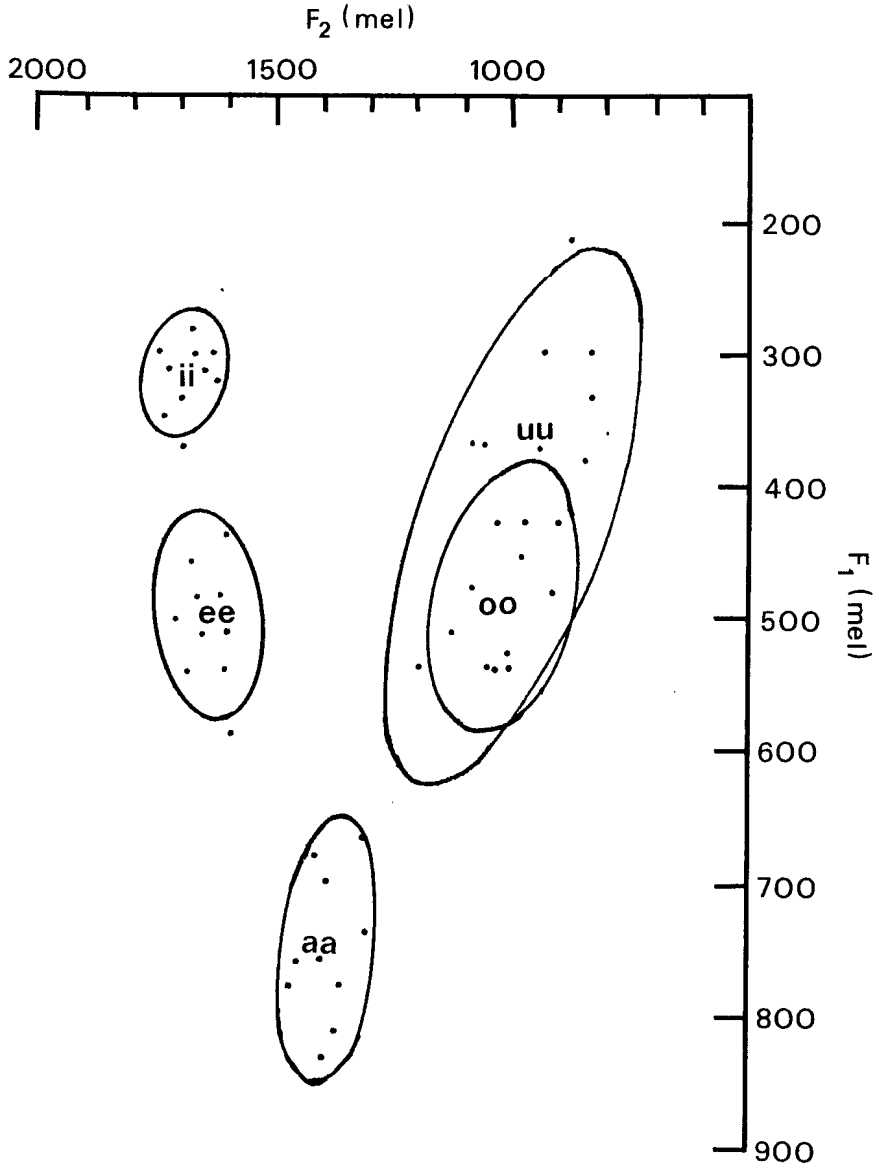


Figure 3. The formant space on a mel scale for the long vowel /ii/, /ee/, /aa/, /oo/, and /uu/ for ten speakers in alveolar environment.

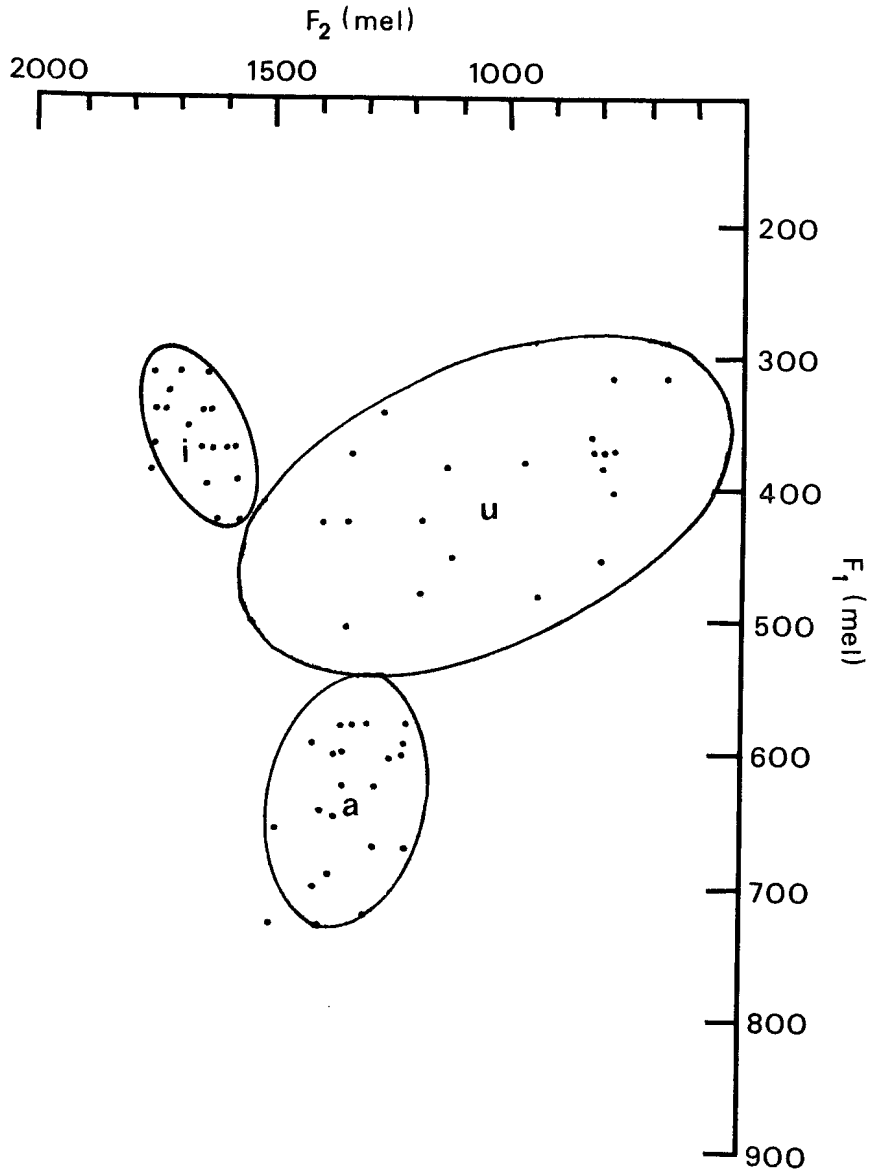


Figure 4. The formant space on a mel scale for the short vowels /i/, /a/, and /u/ in open syllable for ten speakers.

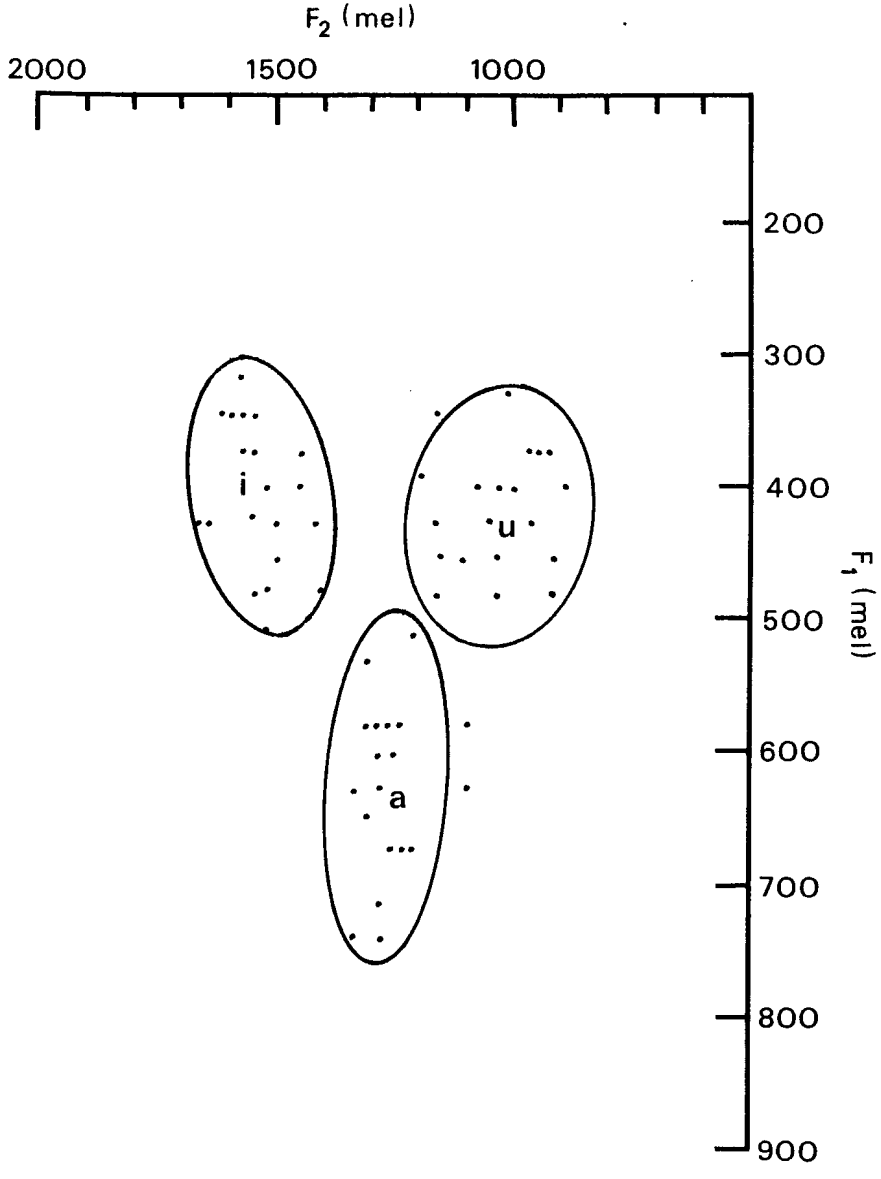


Figure 5. The formant space on a mel scale for the short vowels /i/, /a/, and /u/ in closed syllable for ten speakers.

patterning. In Hausa, only the long vowels in open syllable display the expected durational differences before voiced and voiceless consonants.

The double consonants are geminates. They are considerably longer than the corresponding single consonants. The durational differences are larger between the voiced long and short plosives than between the voiceless pairs. The medial /dd/ is more than twice as long as the /d/, while the /t/ is about two-thirds of the length of the /tt/.

Vowel quality.

Figure 2 shows the F2 - F1 space on a mel scale for the long vowels. The data are from utterances where the long /ii/, /aa/, /uu/ occur after both labial and alveolar consonants, but the mid /ee/ and /oo/ occur after alveolar consonant only. The ellipse of the /oo/ is largely inside that of the long /uu/, but the mean value of the first formant of /uu/ is lower than that of /oo/, placing /uu/ as a higher vowel than /oo/ in this type of vowel space. As can be seeen from the acoustic space in figure 3, however, where the initial consonant environment is restricted to alveolar consonants, the higher extent of the /uu/-vowel is due to the additional labial consonant environment. When the first two formant frequencies of /uu/ and /oo/ after alveolar consonants only are compared by paired t-tests, it turns out that /uu/ and /oo/ do not differ significantly. The frequencies of the third formants were also compared for this vowel pair. As is common for high back vowels in general, the third formant of /uu/ was considerably weaker than for /oo/, and was not visible on spectrograms in some cases. A grouped t-test was used for comparison of the formant frequencies. Their differences were not significant. Kano speakers thus do not distinguish between /uu/ and /oo/ in terms of what we usually call vowel quality. Many of the long /oo/s developed historically from /uu/ in the environment of a following mid vowel (Newman 1979), so it seems almost as if the historical split of /uu/ and /oo/ is about to reverse itself.

The formant spaces for the short vowels are shown in the next two figures. Figure 4 illustrates the short vowels in open syllables, and figure 5 the short vowels in closed syllables, both from environments of alveolar consonants. The three short vowels are well separated in the acoustic space. If short /i/ and short /u/ were realizations of a single high vowel, one would expect to find considerable phonetic overlap when these vowels occur in very similar environments as here. But instead these two high vowels do not overlap in either open or closed syllable. The phonetic evidence thus does not support the notion of neutralization between short /i/ and short /u/. The two short high vowels may be allophones in complementary distribution in many environments (Schachter & Hoffman 1969, Chorier and Faraclas 1981), but in Kano Hausa these two vowels contrast in similar environments. Part of the reason for the large phonetic overlap between these two vowels may lie in the large intrinsic variation of the short /u/. As shown by the size of the ellipse, there is considerably more variation between speakers for the short /u/ than for the other vowels, particularly in open syllable. Thus the short /u/ may be realized by some speakers as quite fronted, but it is still separate from the short /i/. A large variability of high back vowels has been noted for many languages (Ladefoged 1967, Keating 1984). The large amount of variation in the phonetic realization of short /u/ in Hausa is probably more due to this common tendency than to language-specific rules of Hausa.

Vowel duration and vowel quality.

The qualities of the vowels show differences in the three vowel lengths. As in many other languages the long vowels are more peripheral than the short vowels. Paired t-tests between the formant frequencies of the long vowels and the short vowels in open and closed syllables demonstrate that the different vowel lengths, in most cases, are accompanied by different formant frequencies. Long /ii/ and short /i/ in both syllable types all differ significantly ($p<0.001$) along both F1 and F2: the short /i/'s are lower and more back than /ii/. The long /uu/ and the short /u/s all differ significantly ($p<0.001$) along F2, but not along F1. The high back vowels thus differ more importantly in the backness/rounding dimension than in the height dimension. The short low /a/-vowels in open and closed syllables are not significantly different, but the long /aa/ differs significantly ($p<0.001$) from both short /a/s along F1. F2 is quite similar for all the low vowels. The long and short low vowels thus differ in the height dimension.

Thus in similar environments the three classes of vowels tend to differ significantly in duration and quality. The question arises whether the vowel quality differences are predictable from differences in vowel length, or have to be specified independently of length, as separate targets. This question has implications for phonology. If vowel quality differences result from differences in length because of an automatic process of vowel reduction which affects the short vowels and causes them to fall short of the target values achieved by the long vowels, then it argues for specifying Hausa vowels as five basic vowels, and separating out the treatment of length as an independent feature. Vowel reduction can be regarded as an effect of 'undershoot' in the production (Lindblom 1963): the speaker aims at the same target for both long and short vowels, but due to lack of time the short vowels end up short of the intended target. The quality differences between long and short vowels then become results of mechanical constraints of the speech production mechanism. If the vowel quality differences cannot be treated as vowel reduction, then they must be regarded as phonologized, and two or three sets of underlying vowels must be posited, each set specified as to both duration and quality. In this case, each vowel is regarded as being stored with inherent information on quality and timing specified together, in the form advocated in action theory of speech production (Fowler 1980). The question at issue then, is does Hausa have long and short vowels with the same underlying qualities, the differences between the long and short variants being due to mechanical effects, or are the long and short vowel qualities distinctive?

If long and short vowels have the same target (intended quality), then one expects any remaining quality differences between long and short vowels to depend on the consonantal environment. There is a larger articulatory change in going from an alveolar consonant to a back rounded vowel than in going from a labial consonant to a rounded vowel. Conversely, there is a larger articulatory change in going from a labial consonant to a high front [i]-type vowel than in going from an alveolar consonant with a high tongue position to a high front vowel. In other words, the vowel in [di] will be less affected and closer to its target position than the vowel in [du], and the vowel in [bu] will be less affected than the vowel in [bi]. Acoustically, the transition from the low F2 locus of labial consonants to the high F2 of [i]-type vowels involves a much longer distance than the transition to the low F2 of [u]-type vowels. The transition from the relatively high F2 locus of alveolar consonants to the low F2 of [u] is longer than to the high F2 of [i]. An undershoot account, in which it is postulated that corresponding long and short vowels have the same target thus predicts larger

acoustic differences in F2 between /ii/ and /i/ after labial consonants than after alveolar consonants, and larger F2 differences between /uu/ and /u/ after alveolar consonants than after labial consonants. An account that allows a long vowel to have a different target from a short vowel will not make this prediction.

Paired t-tests were applied to the differences in F2 between /ii/ and /i/, and between /uu/ and /u/, in each case after labial and alveolar consonants. Table 2 shows the results.

ii-i	after alveolar consonant	53	(NS)
	after labial consonant	81	($p<0.001$)
uu-u	after alveolar consonant	288	($p<0.001$)
	after labial consonant	50	(NS)

Table 2
F2 differences in mel between long and short high vowels after alveolar and labial consonants.

The results support an undershoot account. The second formant differences are significant when the consonant-to-vowel transition involves a relatively large distance, so that the short vowel ends up short of the intended target. When this transition involves a small distance the vowel quality can be reached in the short vowel as in the long vowel. The quality differences between long and short vowels can thus be accounted for by the mechanics of the speech production system operating on five basic targets. The Hausa vowels are thus best described as a basic five vowel system. These basic vowels have the qualities that appear in phonetically long vowels. The long vowels are derived as double vowels by stringing two of the same basic vowels together, e.g. /ii/. The qualities of the short vowels are derived from the basic vowels by way of applicatin of vowel reduction rules that result in undershoot of the targets.

In addition to the above quality differences resulting from durational differences, Hausa vowel qualities also vary depending on their consonantal environments. Here I just want to mention a few tendencies noted in the data. As is implied by the above discussion, vowels in alveolar environments are further front than the same vowels in a labial environment, and vowels in velar environment tend to occupy a space between the vowels in alveolar and labial environments. Back vowels after laterals are more front than back vowels after other alveolar consonants. The laterals seem to have a fronting effect on back vowels, though not on front vowels. Front vowels are more front after palatal consonants than after other consonants. Low vowels after /h/ are lower than after other consonants.

Diphthongs.

The nature of diphthongs in general is a little studied topic. Most of the units that are labeled diphthongs in the languages of the world are derived from underlying sequences of vowels, or from sequences of a vowel and a glide. Their phonetic realizations can be quite different in different languages. The

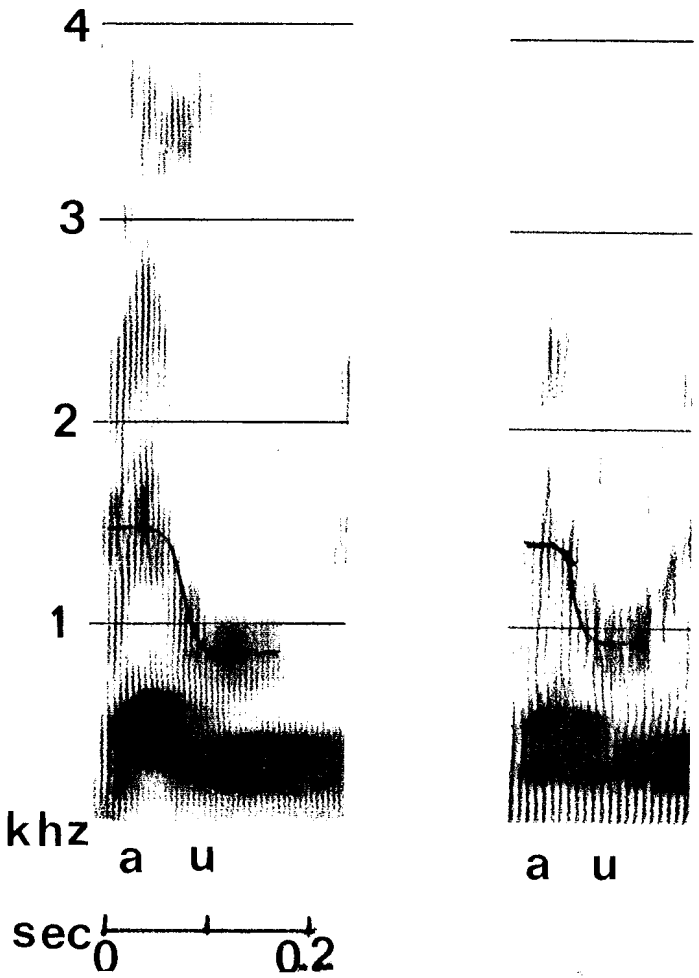


Figure 6. Spectrograms of the /au/ diphthong for two speakers.

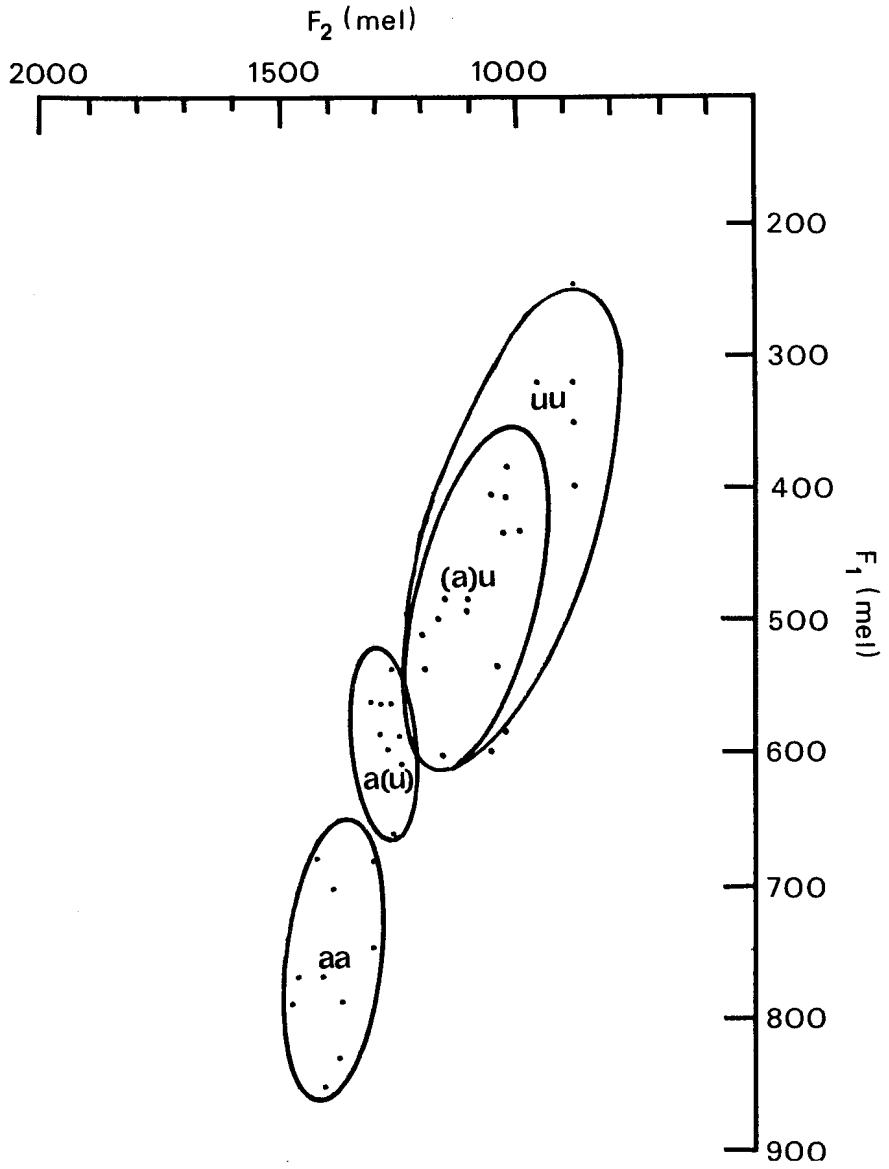


Figure 7. Formant chart of the [a] and [u] parts of /au/, and /aa/ and /uu/ for ten speakers.

relationship between the component parts of diphthongs may vary. In some languages the transition between the vowel components constitutes a major part of the diphthong, in other languages this transition is quite fast, and the major part of the diphthong consists of relatively steady vocalic states surrounding the transition. Consider /ai/ and /au/ in a few languages where data are available. In American English and Peking Chinese the transition occupies a major part in these diphthongs, while in Cairo Arabic, the transition is very fast. In American English the transition provides about 60% of the duration of /ai/ and 75% of /au/ (measurements from data in Gay 1968), and in Chinese it provides about 50% of /ai/ (Svantesson 1984). In Cairo Arabic, on the other hand, the transition occupies only 15% of /ai/ and 25% of /au/ (measurements from data in Norlin 1984). The durations and the relationships between the component parts of a diphthong must thus form part of the phonological rules of a language.

Hausa /au/.

The /au/ in Hausa is always pronounced as a diphthong. Figure 6 illustrates typical examples of /au/ from two speakers. From an acoustic point of view it can be described as two vowels connected by a transition. In addition, there are transitions from and towards surrounding consonants. As these consonant-to-vowel and vowel-to-consonant transitions presumably are not specific to the diphthong, but follow the same rules as for other vowels, they will not be considered in detail here. The questions of concern are: What are the acoustic properties of the vowel components of the diphthong? What kind of model can characterize the acoustic properties of the diphthong?

The vowel qualities of the component parts of the diphthong vary in the same way as other /a/s and /u/s in Hausa depending on the surrounding consonants. Restricting the diphthong to one environment, between alveolar consonants, the vowel qualities of the vowels comprising the diphthong are shown in figure 7. For comparison, /aa/ and /uu/ from the same environment are also included. The onset of /au/ is a low vowel, although it is not as low as /aa/. The quality of the offset vowel approaches that of /uu/. The vowel qualities in /au/ can be derived from the basic /a/ and /u/ in the same way as the qualities of the short vowels by applying vowel reduction rules that result in undershoot of the target qualities due to limitations in duration.

The mean duration of the diphthong in these data is 185 msec. This is considerably longer than the duration for long vowels. The main reason for the diphthong being so much longer here is probably because the diphthong was measured before an alveolar tapped r-sound in zaure 'entrance', and vowels before r-sounds tend to be longer than before plosives. It is also possible that the transition between the vowels in the diphthong adds to its length. The diphthong /au/ can be described as having four parts: a steady state part of /a/, a vowel-to vowel transition, a steady state part of /u/, and a vowel- to consonant transition. The steady state portions of the vowel components occupy roughly 30% each of the diphthong (the mean duration is 55 msec. each). For seven out the ten speakers the transition between the vowels is quite rapid: 20-35 msec. This is about 15% of the diphthong. The remaining three speakers have a considerably longer transition. These three speakers were excluded from consideration here. For most Hausa speakers, however, their /au/ is similar to that in Cairo Arabic, but different from that in American English and Chinese, where the transition constitutes a much larger part of the diphthong. The transition from /u/ to the following alveolar consonant takes up the remaining 25% of the diphthong (its mean duration is 50 msec.).

The spectrograms of the diphthongs also show that F1 and F2 do not always move in synchrony. For most of the speakers the transition part of the diphthong occurs later in F1 than in F2. This F1 delay is about 20 milliseconds. I have also observed a similar F1 lag in spectrograms of American English /au/ for some, but not all speakers. Because the delay in F1 is not apparent in the diphthong in every speaker, it cannot be an inevitable effect of the mechanics of the speech production system. This phenomenon needs further investigation. At this stage I can only speculate that the F1 lag occurs because speakers make use of different timing relationships between lip, jaw, and tongue movements. The F1 lag is best accounted for by optional rules that some speakers make use of and others do not.

The Hausa diphthong can be derived from a string of two vowels. The vowel qualities and durations of this base form are those of the basic /a/ and /u/. Rules of vowel reduction that centralize the vowel qualities (figure 7), duration adjustments and transition insertions generate the phonetic [au]. For an acoustic output the mean formant frequencies of [a] and [u] are taken as input. The mean duration of the steady state of each vowel component is 55 msec. The mean duration of the transition between the vowels is 25 msec. The formant trajectories of the diphthong itself, i.e. excluding the consonant-to-vowel transition and the vowel-to-consonant transition, can be generated from four points, A, B, C, D, illustrated in figure 8. A-B and C-D are the formant

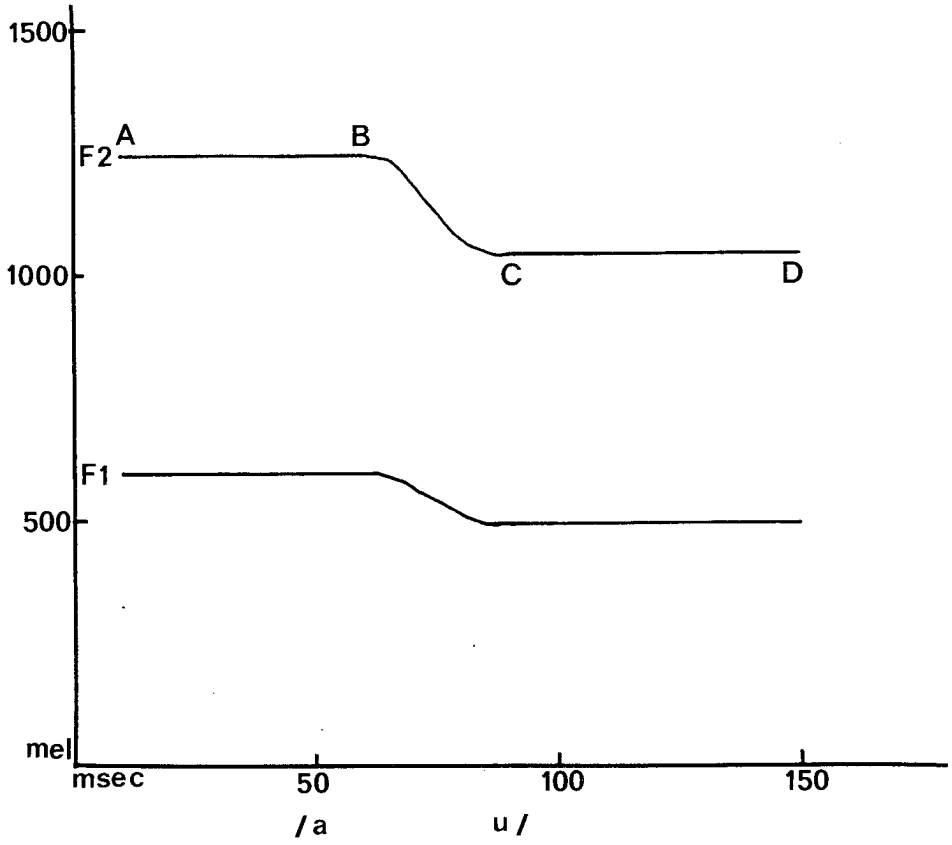


Figure 8. Acoustic model representation of the steady states and the vowel-to vowel transition parts of Hausa /au/, generated by rules discussed in the text.

trajectories of the steady states, each 55 msec. long. The formant transitions between the two vowels do not form just a straight line, but follow an s-shaped trajectory. If the points B and C are considered as (t_1,y_1) and (t_2,y_2), then the interpolation is in accordance with a third order polynomial such that the velocity is zero at each end. Thus the program calculates the constants a_0, a_1, a_2, a_3 in a polynomial of the form:

$$y = a_0 + at + a_2t^2 + a_3t^3$$

by solving the following set of four equations with four unknowns:

$$y_1 = a_0 + a_1t_1 + a_2t_1^2 + a_3t_1^3$$

$$y_2 = a_0 + a_1t_2 + a_2t_2^2 + a_3t_2^3$$

$$\frac{dy_1}{dt} = 0 = a_1 + 2a_2t_1 + 3a_3t_1^2$$

$$\frac{dy_2}{dt} = 0 = a_1 + 2a_2t_2 + 3a_3t_2^2$$

where y = formant frequencies at time t.

Figure 8 is an acoustic representation of the essential parts of an average Hausa [au], generated by the above rules. In addition, there may be a possible, optional delay in the F1 transition. The transition from /u/ to the following consonant is not considered here.

This model of the Hausa /au/ is very general and can be applied to diphthongs in other languages. Diphthongs are represented as two vowels, /VV/, and they are generated by stepping from the first vowel to the second one through a transition. The formant frequencies of the vowel components and their durations are language specific, as is the duration for the transition. As was pointed out above, the transition duration in English and Chinese, for example, generally occupy a much larger part of the diphthong than in Hausa and Arabic. Given the duration of the transition, the formant trajectories can be generated using the equation for a third order polynomial.

Hausa /ai/.

Although the diphthong /ai/ is symbolized as if it involved a considerable change in vowel quality, eight of the ten speakers pronounced /ai/ in labial and alveolar environments as a monophthong [e:]. The two remaining speakers used diphthongal pronunciations; [eI] in alveolar environment and [.I] in labial environment. A diphthongal [.I] pronunciation occurs for most speakers after /w/, as in kwai 'egg'. Figure 9 illustrates typical [e:] realizations of /ai/ on spectrograms of two speakers. Figure 10 is an acoustic chart of the [e:] realization of /ai/, as well as the realizations of /ii/ and /ee/, all from the environment of alveolar consonants. The [e:] realization of /ai/ occupies a higher position on the vowel chart than the [e:] realization of phonemic long

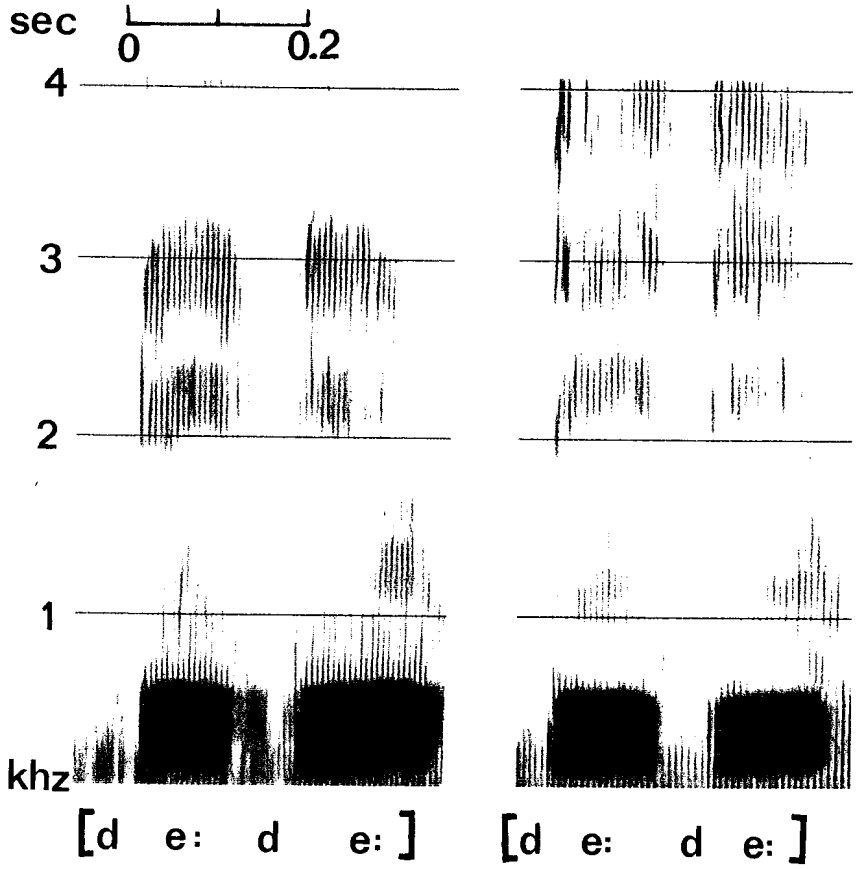


Figure 9. Typical [e:] realizations of /ai/ on spectrograms of two speakers.

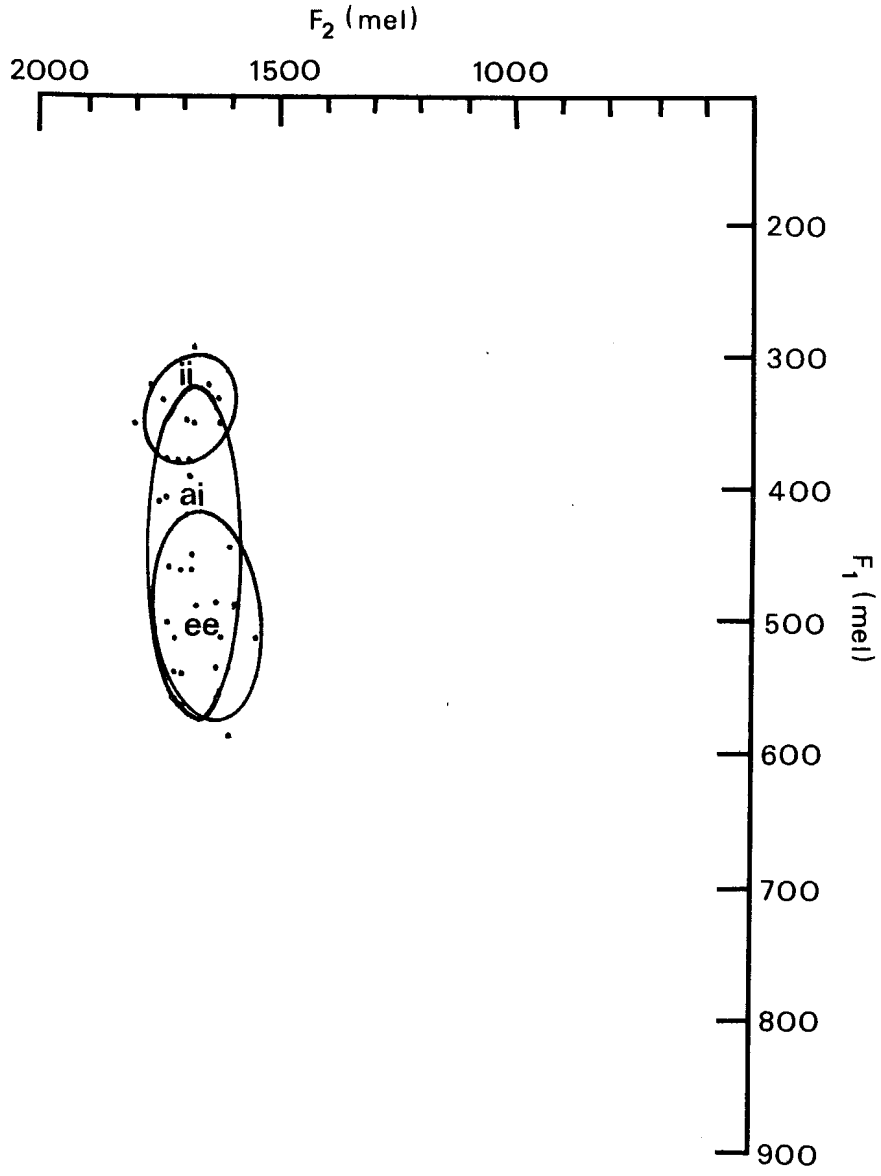


Figure 10. Formant chart of the [e:] realization of /ai/, /ii/ and /ee/ for ten speakers.

/ee/. There is considerable overlap between these two phonetic [e:]'s but their mean formant frequencies are different. These differences are significant for both F1 and F2 (paired t-test; $p<0.005$). It is interesting to notice a speculation by Newman and Salim (1981) that the diphthong /ai/ should perhaps be derived from a long mid vowel that is different from the long /ee/. Newman and Salim symbolize this proposed underlying vowel /EE/. The strong tendency towards a monophthongal pronunciation of /ai/, and the differences in vowel quality between the realizations of /ai/ and /ee/ provide some phonetic support for this proposal by Newman and Salim.

SUMMARY.

Based on phonetic evidence the Hausa vowels are best described as five basic vowels. The long vowels are represented as doubled basic vowels. There is some indication that perhaps the /ai/ diphthong should be derived from a long front mid vowel that is different from /ee/. The contrast between /uu/ and /oo/ seems to be in the process of being lost. The different qualities in the short vowels are derived by vowel reduction processes. The process for generating /au/ consists of stringing underlying vowels together in the same way as for generating long vowels. The results thus give phonetic support to an analysis of Hausa long vowels and the diphthong as /VV/.

References.

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

Chorier, B. and N. Faraclas (1981). A closer look at short high vowels in Hausa. University of California, Berkeley: unpubl. ms.

Fowler, C. (1980). Coarticulation and theories of extrinsic timing. Journal of Phonetics 8:113-133.

Gay, T. (1968). Effect of speaking rate on diphthong formant movements. Journal of the Acoustical Society of America 44.6:1570-1575.

Keating, P. (1984) Vowel allophones and the vowel-formant phonetic space. UCLA Working Papers in Phonetics 59:50-61.

Ladefoged, P. (1967). Three areas of experimental phonetics. London: Oxford University Press.

Lindblom, B. (1963). Spectrographic study of vowel reduction. Journal of the Acoustical Society of America 35:1773-1781.

Newman, P. (1979). The historical development of medial /ee/ and /oo/ in Hausa. Journal of African Languages and Linguistics 1.173-188.

Newman, P. and B.A. Salim (1981). Hausa diphthongs. Lingua 55:101-121.

Norlin, K. (1984). Acoustic analysis of vowels and diphthongs in Cairo Arabic. Lund University, Department of Linguistics Working Papers 27:185-208.

Parsons, F. (1970). Is Hausa a Chadic language? Some problems of comparative phonology. African Language Studies 11:271-288.

Port, R.F., S. Al-Ani, and S. Maeda (1980). Temporal compensation and universal phonetics. Phonetica 37:235-266.

Salim, B.A. (1977). Phonemic vowel neutralization in Hausa. In P. Newman and R.M. Newman (eds.) Papers in Chadic Linguistics, pp.131-141. Leiden: Afrika-Studiecentrum.

Schachter, P. and C.Hoffman (1969). Hausa. In E. Dunstan (ed.) Twelve Nigerian Languages, pp. 73-84. London.

Schuh, R. (1979). Toward a typology of Chadic vowel and tone systems. UCLA, unpublished ms.

Svantesson, J.-O. (1984). Vowels and diphthongs in Standard Chinese. Lund University, Department of Linguistics. Working Papers 27:209-236.

Acknowledgments: I would like to thank Will Leben, Brian McHugh, and Nicolas Faraclas for taping all the speakers in Kano. I would also like to thank Paul Newman, and Peter Ladefoged, Ian Maddieson, Michel Jackson and the rest of the Phonetics Lab at UCLA for the many helpful comments and discussions. This work was funded by the National Science Foundation and the Swedish Research Council for the Humanities and Social Sciences.

The registers of Mon vs. the spectrographist's tones.

Gérard Diffloth

University of Chicago

In a recent issue of WPP, Thomas Lee presented the results of his acoustic research on Mon registers, and concluded: "our study does not support Diffloth's claim that the fundamental difference between the two Mon registers lies in a phonation difference. ... The most significant parameter of the register distinction is that of pitch. Indeed as Shorto (1962) suggests, Mon is a quasi-tonal language" (Lee, 1983: 94-5).

Lee's conclusion will surely come as a surprise to linguists acquainted with the Mon language: everyone agrees that it is a true register language, i.e. one where phonation type distinctions (in this case, clear voice vs. breathy voice) are contrastive; and Shorto is in fact the foremost exponent of this notion.

I have to admit that I am responsible, to a great extent, for Lee's misinterpretation of Shorto's views: in a recent article (Diffloth, 1982), I tried to extract from Shorto's use of the term "quasi-tonal" more than he apparently had intended to say, namely that pitch might be an element in the registers of Mon[1].

In another publication, Shorto talked of the Mon "para-tonal register distinction" saying: "its exponents are distributed throughout the articulatory complex but exclude pitch features" (Shorto, 1966: 399-400). and in 1967, Shorto dropped the troublesome terms "quasi-tonal" and "para-tonal" altogether, concluding: "Pitch differences as an exponent of register is lacking" (Shorto, 1967: 246).

In view of experimental findings concerning the mechanism of breathy voice production, Shorto's statements on the question of pitch appear to be possibly a little extreme: it is easier for the vocal apparatus to use the lower part of the pitch-range in order to produce a breathy voice, than the higher part. As John Laver has cautiously written "High pitched breathy voices seem rare" (Laver, 1980: 133).

I have pointed out (Diffloth, 1982) that in word-by-word elicitation, clear-voice Mon syllables tended to have a higher pitch than breathy-voice ones, but that in the normal flow of speech, while phonation types remained clearly distinct, intonation completely superceded the pitch effect. The last part of this remark was possibly itself a little extreme, and I will gladly correct it, if faced with experimental evidence to the contrary[2].

But the point on which everyone but Lee agrees is that, for Mon, "head register is characterized by clear voice, ...Chest register....by a breathy voie" and "contrastive voice quality is always present, and is probably the feature most readily perceived" (Shorto, 1967: 246). See also: (Haswell, 1874; Blagden, 1910, Halliday, 1922; Huffman, 1976, 1977; Sakamoto, 1974; Diffloth, 1982, 1983, 1984). So much for the record, now to the point itself.

Lee made wide and narrow band spectrograms from recordings of Mon words, in order to measure vowel duration, formant frequencies and fundamental frequencies. His purpose was to explore the acoustical parameters of the register distinction

and their relative significance. Much of the paper centers on the problem of the relative importance of pitch vs. phonation types, and so does the conclusion.

But it should have been clear from the start that pitch and phonation type are not commensurable entities: what numerical acoustic coefficient could possibly tell us that one is more important than the other, let alone more significant? An answer to Lee's question does not come from acoustical measurements alone; it would have required the use of a speech synthesizer able to imitate a wide spectrum of phonation types, as well as pitches, and the computation of recognition and error responses from native speakers of Mon.

What Lee attempted to do instead was to deny the significance of Mon phonation types altogether, by showing that in several cases (two vowels, to be precise) he could not detect any such distinction on the spectrograms he made.

In order to detect possible breathy phonations[3], Lee measured amplitude differences between F_0 and other formants at one third of the duration of the vowel, a measurement which has been shown to be relevant to the phonation types of Gujarati and !Xõo. He found, after averaging out five distinct regional dialects of Mon, that the vowel /o/ did actually show a contrast in phonation types, but not /a/ nor /./. This result, like his general conclusion, does not reflect anything anyone else has ever observed in studying Mon. Therefore, one wonders about the attainable precision and the inherent relevance of this approach to phonation types.

As for the other vowels of Mon, Lee decided to exclude them from his test, on the grounds that there were small vowel quality differences between the clear and breathy members of each pair; this would disturb the measurement of formants (Lee, 1983: 94). But as Shorto (1966) and many others have pointed out, these vowel quality differences are a typical feature of the sound system of Mon: every modern dialect has its own pattern of vowel warping. It is actually unusual to find pairs of breathy vs. clear vowels with absolutely identical vowel qualities. If Lee's technique requires identical formant frequencies, this limitation prevents him from studying Mon phonation types in any meaningful way.

In fact, most Mon Khmer languages which are reported to have a phonation type distinction also show slight or sizable vowel quality differences between the members of each vowel pair (see e.g. Thongkum, 1979). Lee's technique would not apply to any of these languages either.

More generally, the sound spectrograph is well suited for measuring fundamental frequencies, formant frequencies and amplitudes; but it was not built for studying phonation types. In order to obtain, for breathy and clear voice, measurements of a precision and relevance similar to those Lee obtained for pitch, a direct experimental study of the laryngeal activity of Mon speakers would have been necessary[4].

Besides, I do not see the point in making a statistical salad out of regional dialects purposefully chosen for their systematic differences.

Lee has not used the instrumental techniques necessary for measuring differences in Mon phonation types, differences which the trained ear can easily perceive, and the trained larynx acceptably produce.

Footnotes

1) It was not clear, in the footnote to the relevant passage (Shorto, 1962:x.) whether his remark "the exponents of register.... exclude pitch features" applies to Cambodian alone or includes Mon as well.

2) Nearly all the data used by Lee consist of word citations. Only three short sentences are used; one of them "say the word....once again" contains a quotation intonation in its relevant portion; the other two: "give me some...", "give me a...", two tokens each, were recorded out of any pragmatic context: for intonation purposes, they should be described as sentence-citations. Connected speech it is, but surely not the "normal flow of speech", as far as intonation is concerned.

3) In this discussion, I have assumed that we are not simply quibbling over terminology. The term "breathy voice" has been used to refer to several distinct phonation types which are auditorily similar. I am using the term in a wide sense, considering that "incomplete closure of the vocal folds" is "the main characteristic of breathy voice" (Laver, 1980: 146). Lee himself does not allude to this question.

4) The Phonetics Laboratory at Chulalongkorn University, Bangkok, is undertaining experiments with native speakers of Mon and several other Mon-Khmer languages, using fiberoptic and glottographic equipment. T.L. Thongkum is the head of that project. I will be responsible for providing the comparative background.

Editor's note

Since the recordings analyzed in Lee's 1983 paper were obtained through the assistance of Gérard Diffloth we especially welcome the opportunity to include this dissenting view of the data in UCLA WPP. In fairness to Lee it might be pointed out that the information that the recording included "regional dialects purposely chosen for their systematic differences" did not reach him until after the analysis was completed. It might also be added that, although some other technique might reveal a difference in laryngeal setting for the registers in Mon, careful listening by several persons with "trained ears" in the UCLA Phonetics Laboratory does not suggest that breathiness is at all a consistent feature of "chest register" in the tape recordings we have, whereas an observable and statistically reliable pitch height difference does occur.

References

Blagden, C.O. (1910) "Quelques notions sur la phonétique du Talain et son évolution historique" Journal Asiatique 477-505.

Diffloth, G. (1982) "Proto-Mon registers: two, three, four...?" Proceedings of the Eighth annual meeting of the Berkeley Linguistic Society, University of California Berkeley.

Diffloth, G. (1983) "Registres, dévoisement, timbres vocaliques: leur histoire en Katouique" Mon-Khmer Studies XI 47-82.

Diffloth, G. (1984) The Dvaravati-Old-Mon language and Nuah-Kur. Monic Language Studies, Vol. I. Chulalongkorn University Press, Bangkok

Halliday, R. (1922) A Mon-English dictionary. Siam Society, Bangkok. Reprinted in 1955 by the Mon Cultural Section, Ministry of Union Culture, Government of the Union of Burma, Rangoon.

Haswell, J.M. (1874) Grammatical notes and vocabulary of the Peguan language. American Baptist Press, Rangoon; 2nd Ed. 1901.

Huffman, F.E. (1976) "The register problem in 15 Mon-Khmer languaages" Austroasiatic Studies. Pacific Linguistics, Special Publication No. 13, Hawaii.

Huffman, F.E. (1977) "An examination of lexical correspondances between Vietnamese and some other Austroasiatic languages". Lingua 43: 171-98.

Laver, J. (1980), The phonetic description of voice quality. Cambridge Studies in Linguistics, 31. Cambridge University Press, Cambridge.

Lee, T. (1983) "An acoustical study of the register distinction in Mon" UCLA Working Papers in Phonetics 57: 79-96.

Sakamoto, Y. (1974) "Notes on the phonology of Modern Mon (Prakret Dialect)" Journal of Asian and African Studies, Tokyo. (In Japanese).

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

Shorto, H.L. (1966) "The Mon vowel system, a problem in phonological statement" in: In memory of J.R. Firth, ed. by C.E.Bazell et al., Longmans, London.

Shorto, H.L. (1967) "The register distinction in Mon-Khmer languages" Wissenschaftliche Zeitschrift der Karl Marx Universität Leipzig 16: 245-8.

Thongkum, T.L. (1979) "The distribution of the sounds of Bru" Mon-Khmer Studies VIII 221-93.

"Tense" and "lax" in four minority languages of China

Ian Maddieson and Peter Ladefoged

1. Introduction

Most sets of proposed phonological features include a feature [tense/lax] (e.g. Jakobson, Fant and Halle 1952, Chomsky and Halle 1968, Stevens, Keyser and Kawasaki 1984). It is obviously convenient to have a label for phonological groupings of vowels that behave in a similar way. For example, in Standard British English the stressed vowels /ɪ, ɛ, æ, ʌ, ɔ, ɑ/ are restricted to closed syllables, and these are the only vowels which can appear before /ŋ/. This set of vowels is often labeled "lax", in opposition to the rest of the vowels, which can occur in open syllables and are labeled "tense". Similar groupings of vowels in other Germanic languages are labelled similarly (Linell et al. 1971, Jørgensen 1969, Spa 1970). But is this grouping of vowels phonologically natural? And if so, is there a phonetic basis to this classification such as to justify including a feature [tense/lax] in the set of universal phonological features?

The description of vowels as "tense" or "lax" has a considerable history behind it. The claim that vowels differ in the degree of muscular tension in the tongue seems to originate with Bell (1867), who used the terms "primary" and "wide". It was a basic part of the phonetic theories of Henry Sweet from his Handbook of Phonetics (1877) on, and of Jespersen (1899), who may have been the first to use the terms "tense" and "lax" for its description. Sweet gave several accounts of the distinction, which he called "narrow" vs "wide". One of them reads as follows:

> in a narrow vowel the tongue is bunched or made convex lengthways, and there is a feeling of tension or clenching; in wide vowels the tongue is relaxed and comparatively flattened. (Sweet 1908).

In Sweet's view the tension parameter is important not only for grouping vowels within languages, but also for characterizing cross-language differences. However, the validity of the parameter has never been demonstrated. Daniel Jones, Sweet's principal successor in the British phonetic tradition, rejected the distinction in the following passage:

> Tense vowels are those which are supposed to require considerable muscular tension on the part of the tongue; lax vowels are those in which the tongue is supposed to be held loosely. .... It is not by any means certain that this mode of describing the sounds really corresponds to the facts. (Jones 1956: 39).

Jones points out that the English lax and tense vowels are distinguished by their relative lengths and that such pairs as /i / and /ɪ / can be described in terms of different positions on height and backness parameters that are universally recognized. Parallel accounts of the other Germanic languages can also be found (Elert 1964, Wängler 1958, Pols 1977, Fischer-Jørgensen 1972). If the classificatory feature is length, and the quality of individual vowels can be described by standard parameters, then there is no need for a feature [tense]. (Perkell (1969) attempts to account for the durational differences in terms of articulatory strength, stronger contractions predicting longer steady states for tense vowels.)

A different characterization of the tense/lax difference, first proposed in a 1966 paper by Stockwell (Stockwell 1973), suggests that tense vowels differ

from lax vowels by being closer to the periphery of the vowel space than lax vowels. Chomsky and Halle's (1968) definition of their [tense/lax] feature includes the claim that tense vowels "are executed with a greater deviation from the neutral or rest position of the vocal tract". Lindau (1978), following Stockwell's suggestion, argues that this property of tense vowels should be isolated from questions of muscular tension (and of pharyngeal volume, see below) and that the feature should be renamed [peripheral]. However, as Catford (1977) had already pointed out, the English lax vowels do not all lie closer than the tense vowels to a neutral vowel position or further from the periphery of the traditional pseudo-articulatory vowel space. In particular, /æ/ or /ɔ/ are not less peripheral than, say, /ɑ:/. These difficulties become more severe if the vowels are examined in an acoustic space (Disner 1983, Lindau 1978) or if the eccentric assumption that the neutral vowel position is /ɛ/ is made (Chomsky and Halle 1968). And, in any case, peripherality can be expressed in terms of height and backness parameters and hence it is not a phonetic prime.

Another attempt to rescue tenseness as a phonetic feature derives from Perkell (1971). He suggests that tense vowels have an advanced tongue root, creating an expanded pharyngeal cavity. This idea may have been prompted by the knowledge that the feature which distinguishes vowel harmony sets in African languages with vowel harmony is pharyngeal cavity size (Ladefoged 1964). In languages such as Akan, Igbo and Lango the vowel set with advanced tongue root has often been referred to as "tense" (e.g. Schachter and Fromkin 1968). Perkell's proposal would provide a unified account of the facts of both the Germanic languages and the Niger-Kordofanian and Nilo-Saharan languages of Africa. However pharyngeal width is partly the result of the positioning of the body of the tongue. Lindau showed that in general the pharynx width could be predicted from the tongue height in the radiographic vowel data from German and English available to her. On the other hand in Akan, for example, pharynx width is uncorrelated with tongue height. It follows from this set of observations that a feature [Expanded Pharynx] needs to be recognized for languages like Akan, but that this feature is not involved in the Germanic contrast of "tense" and "lax" vowels.

In a third area of the world, Southeast Asia, the established use of the terms "tense" and "lax" suggests that they describe principally some aspects of the laryngeal setting. In fact, in certain respects scholars of South East Asian languages reverse the way the terms are used in talking about Germanic languages, using the term lax to denote vowels that are both longer and higher in vowel quality than their tense counterparts. These South East Asian vowels are regarded as lax and tense because of other phonetic properties, such as the actions of the vocal cords. In the view of many linguists working on these languages (e.g. Henderson 1967, Gregerson 1976, Matisoff 1973, Egerod 1971, Thurgood 1980, Wheatley 1982), there are certain "normal" associations between tenser and laxer settings of the vocal cords and other aspects of vowel quality, pitch height and contour and even some properties of adjoining consonants. Matisoff, for example, groups clusters of properties into a "tense-larynx syndrome" and a "lax-larynx syndrome". We will show with detailed phonetic evidence that these associations between phonatory and non-phonatory phonetic features in lax versus tense contrasts can differ quite markedly from language to language in the same linguistic area. We will then discuss the implications that this has for recognition of phonological categories of tense and lax vowels.

Specifically, we will explore the meaning of the terms tense and lax in descriptions of four non-Chinese languages spoken in Southern China: Jingpho,

Hani, Yi (Nasu), and Wa. We will do so through an examination of acoustic and aerodynamic properties of the vowels, using data obtained during a trip to Peking and Kunming. By the standards we usually try to follow, the data are very limited in that they were obtained from only one to three speakers of each language. We would have liked to have been able to record six to ten speakers of each language, so that we could be sure that we could report data on the properties of the languages, and not what are possibly just characteristics of individual speakers. But, despite this shortcoming, we thought that we should make our results available because there is so little previously published phonetic data on these languages.

2. Procedure

A number of different techniques were used in the collection of data from these languages. Two physiological variables were recorded: the pressure of the air in the mouth and the rate of airflow out of the mouth. The oral pressure was recorded by means of a small transducer at the end of a thin (3mm) tube, which was inserted between the lips. The procedure was satisfactory only for words containing bilabial consonants. When pronouncing words containing other kinds of consonants, the tube either interfered with the articulation (in the case of dental and alveolar consonants), or did not record the oral pressure behind the articulation (in the case of consonants articulated behind the alveolar region). It would have been better if it had been possible to insert the tube through the nose, so that the oral pressure could have been sensed by means of a transducer in the pharynx, as is discussed by Ladefoged and Traill (1984). But this was not possible in the case of these speakers.

The oral airflow (together with any concurrent nasal airflow that might have been present) was recorded by means of a specially designed and calibrated face mask, similar to that described by Rothenberg (1973). The frequency response of this system is flat (+ 3 dB) up to 2,000 Hz. The airflow and oral pressure records were reproduced in the field on an ultra-violet strip-chart recorder that has a flat frequency response from 0-1000 Hz. The major advantage of this system is that, since these instrumental records were available for inspection during the recording sessions, the experimenter (P.L.) was able to form new hypotheses and record additional material, if appropriate, while the speakers were still available. Furthermore, as the speakers recorded were often either professional linguists or students of their own languages, they were themselves pleased to receive copies of the instrumental records, and were motivated to think of extra items that could be tested. A sample of the output showing a pair of contrasting words in Wa is given as Figure 1. In this figure and throughout the paper the tense vowels are marked by a subscript tilde, e.g. a̰. The sessions during which these aerodynamic records were obtained were tape-recorded. A high quality audio recording of the data was also obtained without use of the mask.

As a first approximation, we may consider the airflow through the mask during a vowel as being equivalent to the airflow through the glottis at that time (during the release of a consonant closure this is not true, but all the measurements to be reported were made during the steady state part of the vowel). As has been shown by Ladefoged (1967), during a voiceless stop in which the vocal cords are apart, the peak pressure of the air in the mouth is a good indication of the subglottal pressure at that time. We are on less certain ground when we infer the subglottal pressure during the vowel from this measurement, but there is no reason to believe that it is substantially different at a point about one third of the way through the vowel, the point at which the flow measurements were made.

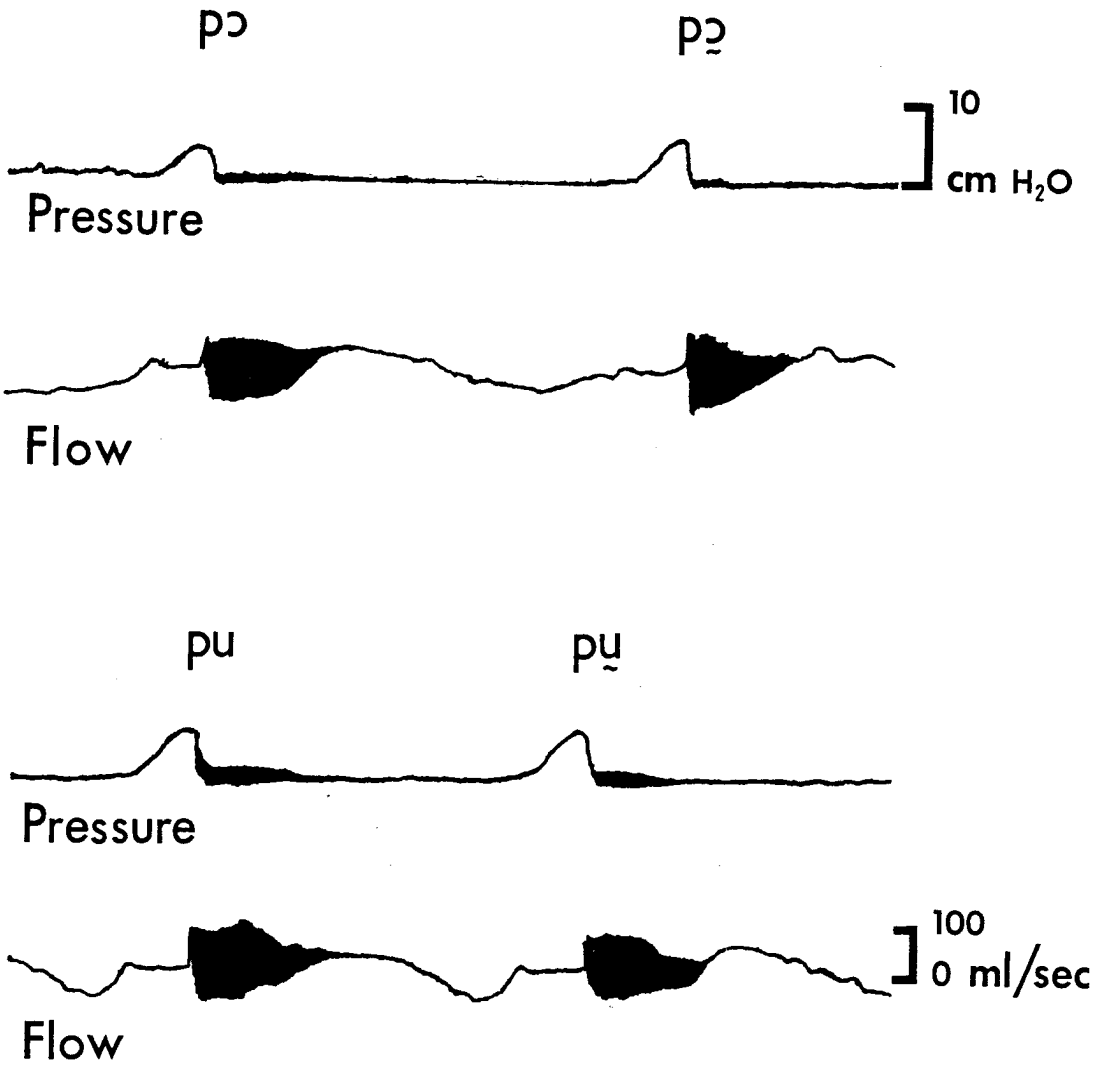


Figure 1. Airflow and intraoral pressure records of two pairs of words with lax and tense syllables in Wa, retraced from the ultraviolet recordings (see Table 6).

The aerodynamic data were used to infer the state of the glottis. In comparing two productions of a vowel with the same vocal tract shape, for any given subglottal pressure the airflow will be greater when the vocal cords are vibrating in the comparatively lax way that produces a slightly breathy phonation type; conversely the airflow will be less during a more tense, slightly laryngealized, or creaky voiced vowel. The subglottal pressure will also affect the rate of airflow; accordingly, if we are trying to use aerodynamic data to determine the glottal state, we must consider the ratio of airflow and pressure. When the ratio of airflow to pressure in an otherwise unchanged vowel is high we will conclude that the vibrating mass of the vocal cords is relatively slack or the vocal cords are less closely approximated; and when it is low, that the vocal cords are relatively stiff or more closely approximated.

The audio record was used for acoustic analyses. The fundamental frequency (F_o) and the frequencies of the first three formants (F_1, F_2, F_3) were measured from wide and narrow band sound spectrograms as appropriate. Measurements were also made of the duration of the vowels, defined as the interval between the onset and offset of detectible (quasi-)periodic pulses in the second formant of the vowel, and hence excluding initial aspiration, final devoiced portions or very irregular final pulses suggesting extreme glottal stricture (i.e. a glottal stop). In addition, certain properties of the initial consonants were measured, such as the voice onset time (VOT) for voiceless stops, and the duration of voicing in voiced consonants. The spectrograms were further inspected for other qualitative differences that might be observable in syllables of the tense and lax classes in each language.

The power spectra of the vowels were obtained and the relative amplitudes of the fundamental and the second harmonic were measured. It has been shown by various authors that in a breathy voice there is comparatively more energy in the fundamental and less in the higher harmonics, whereas in a vowel pronounced with a more constricted glottis the reverse is true (Ladefoged 1981, Bickley 1982, Ladefoged 1983, Kirk et al. 1984). Moreover, variations in the relative amplitudes of the fundamental and the second harmonic measure correlate quite well with listeners' judgments on degrees of breathiness (Ladefoged and Antoñanzas 1984), and reflect differences in "spectral tilt" generated in models of the voice source by varying the rate of vocal cord closure in the glottal pulse (Fant 1983).

3. Data Analysis.

Our descriptions will assume that the difference between tense and lax syllables in the languages under analysis is a property of the vowel, rather than of the tones or the consonants, or the whole syllables. Various different phonological interpretations are possible; but, as we will see, the phonetic differences do seem to be mainly in the vowels. We will, however, talk both of lax and tense vowels and of lax and tense syllables without at this point implying a preference for one phonological analysis over another. We will consider the data for each of the four languages separately.

(a) Jingpho

Jingpho is a Tibeto-Burman language spoken in Southern China and Burma. The two speakers for this study came from the Southeastern part of Yunnan Province, China, from villages fairly close to the Burmese border. They both claimed that their language was the same as that spoken in the Kachin Province of Burma, a

form of Jingpho that has been described by Maran (1971). Kachin Jingpho has no tense/lax vowel contrast. However, the Jingpho grammar published by the Chinese Academy of Sciences (Anonymous 1959) describes a dialect which is quite distinct from Maran's. According to the analysis presented in the Chinese grammar and accepted by our subjects, Yunnanese Jingpho may be considered to have ten vowels, a set of five lax vowels /i,e,a,o,u/ and a set of five tense vowels /i̱,e̱,a̱,o̱,u̱/. There are also three tones, a high tone (55 or 5 in the Chao (1930) notation), a mid tone (33), and a low or low falling tone (11/31). The mid tone does not occur in checked (i.e. stop-final) syllables.

The acoustic parameters of Jingpho were measured for both speakers from the set of words shown in Table 2. The words were spoken in isolation.

Table 2
Words used for acoustic measures on lax and tense vowels for two speakers of Jingpho.

lax		tense	
ti 55	"chubby"	ti̱ 55	"shut one's eyes"
ke? 5	"solidify"	ke̱? 5	"girth"
ka 33	"work"	ka̱ 33	"write"
ko 33	"lay bricks"	ko̱ 33	"tusk"
tu 33	"officer"	tu̱ 33	"grow"

There were no reliable observable differences in the formant frequencies in the two sets, leading us to conclude that Jingpho lax and tense vowels do not differ in quality, unlike the lax and tense vowels of the Germanic and African languages mentioned above. There were also no consistent differences in the duration of lax and tense vowels. For speaker 1 lax vowels had a mean duration of 583 msec, tense vowels 589 msec; for speaker 2 lax vowels were 346 msec, tense 331 msec. (There was, however, a difference between the vowels in open syllables and in syllables checked by a glottal stop, which were considerably shorter - being only on the order of 170 msec in length). The most consistent acoustic difference in the vowels was in the relative energy in different parts of the spectrum. Both speakers pronounced the lax vowels in Table 1 with relatively more energy in the fundamental than in the second harmonic in comparison with the paired tense vowels, with one exception in which there was no difference. Across the two speakers, the fundamental was 3.6 dB higher than the second harmonic in lax vowels but 4.0 dB lower in tense vowels, a difference that is highly significant ($p < .001$). As noted in the previous section, this difference correlates well with the perceptual and productive difference between more breathy and more constricted phonation. The spectra of a tense/lax minimal pair of Jingpho vowels are illustrated in Figure 2. It may be seen that the difference in the relative amplitude of the harmonics between this pair of vowels is very subtle.

In addition to the vocalic differences, there was also a small but noticeable difference in the consonants. In initial stop consonants the VOT tends to be longer before a lax vowel, invariably so for speaker 1 who had a mean VOT of 35 msec in lax syllables, but only 15 msec in tense syllables. Speaker 2 showed a less consistent difference but had no pairs in which the VOT was longer

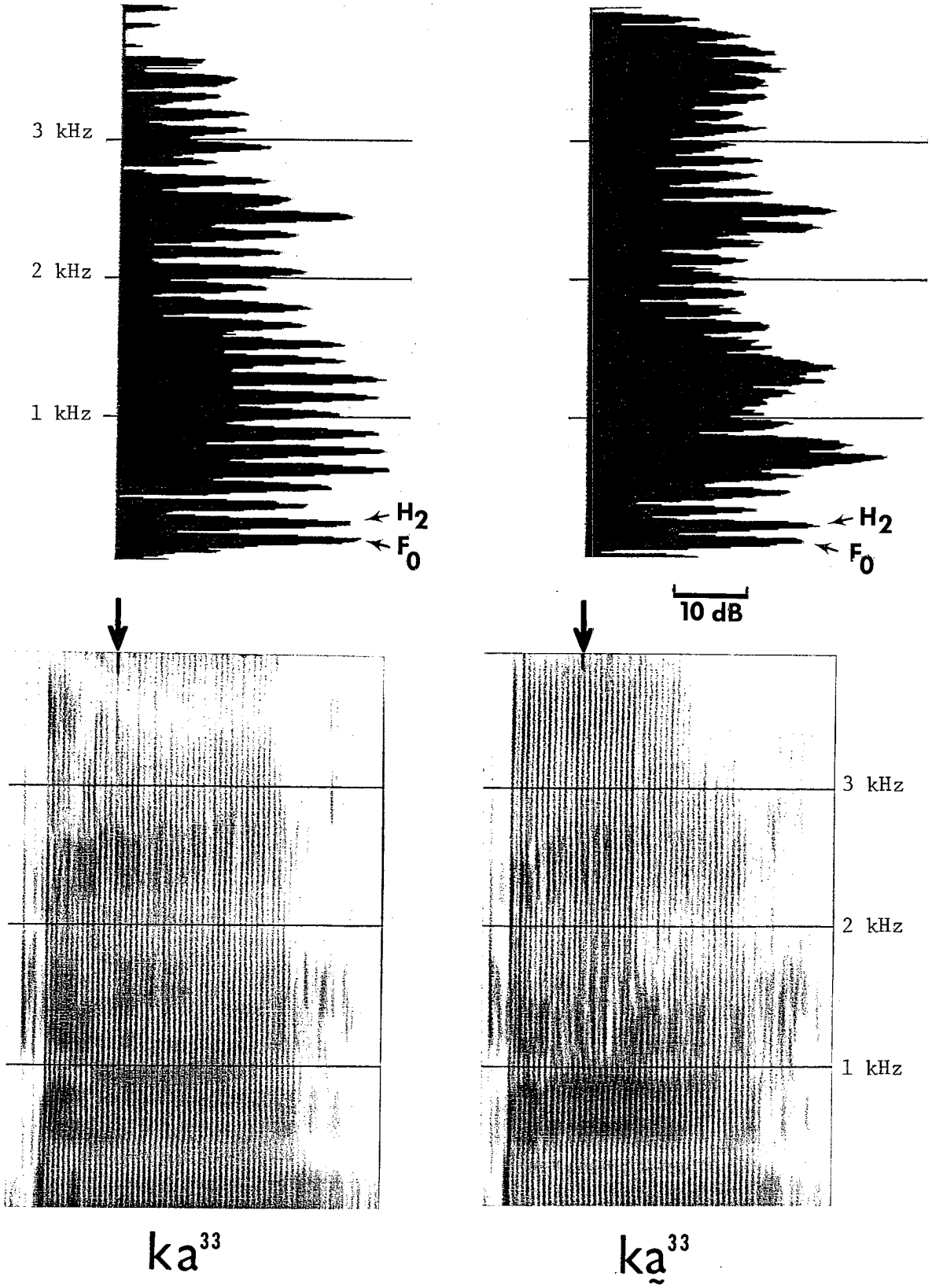


Figure 2. Wideband spectrograms and spectral displays of a pair of mid-tone words with lax and tense syllables in Jingpho. The spectrum is of a 22 msec window preceding the point marked in the spectrograms (see Table 2.)

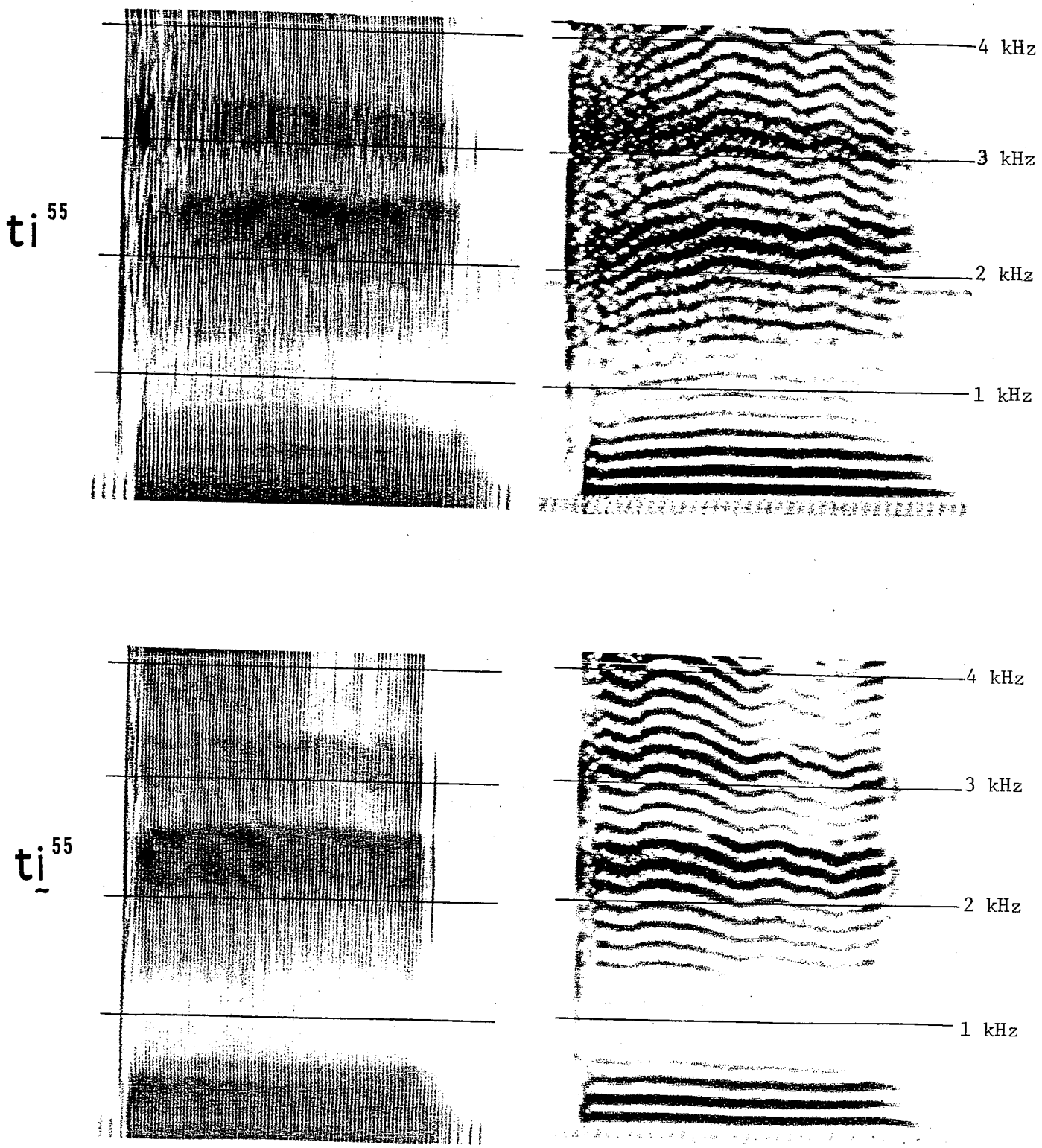


Figure 3. Wide and narrow band spectrograms illustrating the VOT and F_o differences in high-tone syllables in Jingpho (see Table 2).

in the tense member of the pair. Along with this difference in the consonants, one marked difference in F_o was observed. In high tone syllables with lax vowels the onset value of F_o is relatively low and there is a rising pitch contour after the initial consonant, whereas with tense vowels pitch tends to fall from a relatively high onset value. No comparable effect was observed in the mid tone syllables examined. Wide and narrow band spectrograms illustrating the VOT differences and the distinctive F_o contours on lax and tense high tone syllables are given in Figure 3. These spectrograms also make it plain that there is no difference in the vowel quality in this pair of words.

From a physiological point of view, the main difference between the Jingpho lax and tense vowels appears to be in the state of the glottis. A difference in phonation type may be inferred from records of the airflow and pressure. We infer that the one speaker of Jingpho for whom we have both flow and pressure records did not differentiate between tense and lax vowels by using greater respiratory effort for the tense syllables since, in a set of mid tone nonsense syllables containing each of the five pairs of vowels, there were no significant differences in the pressures associated with the two classes of vowels. There were, however, reliable differences in the airflow (and, hence, in the ratio of flow to pressure). Because we know that the formant frequencies are the same in the two vowel sets, indicating no change in supraglottal resistance to airflow, we conclude that the vocal cords are not open as much in tense syllables. This difference in flow was also found for our second speaker. As shown in Table 3, these differences were greater (and statistically more significant) in open syllables than in syllables closed by a glottal stop.

Table 3
Maximum airflow in the vowels in Jingpho tense and lax syllables in ml/sec. n = the number of measurable tokens summed for the two subjects. The statistical significance (p) of the difference is also shown.

	Lax	Tense
CV	112	86
	n = 14	n = 14
	$p < .001$	
CV?	125	105
	n = 11	n = 11
	$p < .02$	

Table 3 also shows that there is a greater mean flow in both lax and tense syllables closed by a glottal stop than in the corresponding lax or tense open syllables. The acoustic records of these words indicate that they are pronounced with greater intensity. Since both lax and tense words are pronounced in this way we conclude that words ending in a glottal stop were produced with greater respiratory force. We do not know why words closed by a glottal stop should be pronounced in this way; perhaps each syllable has a certain quantum of energy assigned for its production, and the shorter the syllable the greater the amplitude of flow must be.

(b) Hani

The "Hani" dialects belong to the Loloish branch of the Tibeto-Burman family (Wheatley 1982). The best known of the Hani dialects, Lüchun Hani, is essentially

the same as the dialect known as Akha in Burma and Thailand (Hansson 1984). We were able to record data of the much less well-known Haoni dialect spoken in Mojiang and Yuanjiang counties in the Yuxi region. The distribution of tense and lax syllables is much more limited in this dialect than it is in other Hani dialects (Hu and Dai 1964) or in Yunnanese Jingpho. Haoni may be considered to have 15 vowels, eleven of them being lax and four being tense, with the tense mid vowel /ɛ̰/ being rare (Li 1979). The language may therefore be considered to have four pairs of vowels in tense/lax contrast with qualities /ḭ/, /ɿ̰/, /ṵ/ and /ɛ̰/. (/ɿ̰/ is an "apical" vowel. The vowel written here with /ṵ/ is analyzed in other sources as a syllabic labial fricative and written /ɣ̰/. No friction is evident in the pronunciation of our speaker.) There are four tones in Hani; high (55), mid (33), and low falling (31), with the fourth tone (35) occurring only on loan words from Chinese. The tense-lax opposition occurs only on the mid and low falling tones. Vowels in high and rising tone syllables are lax. Illustrative words are given in Table 4.

Table 4
Words illustrating lax and tense vowels in Hani (Haoni dialect).

lax			tense		
zɿ	31	"walk"	zɿ̰	31	"(tree, sp.)"
tsɬ	33	"swollen"	tsɬ̰	33	"itch"
ti	31	"beat"	tḭ	31	"live"
tɕi	33	"not at ease"	tɕḭ	33	"scratch"
xɛ	31	"plough"	xɛ̰	31	"eight"
tu	31	"dig"	tṵ	31	"fire"
pu	33	"blow (wind)"	pṵ	33	"sated"

Somewhat limited data, consisting of a number of repetitions of the seven pairs of words in Table 4 were recorded by a single speaker. Acoustically speaking, there is a consistent difference in the distribution of energy in the spectrum. In the seven pairs above, the fundamental is a mean of 2.4 dB higher than the second harmonic in lax vowels, but 2.6 dB lower in tense vowels. This difference is significant at better than the .01 level. Thus the lax vowels may be said to be more breathy than the tense vowels. However, it appears that the difference between tense and lax syllables is more complicated in Haoni than in Jingpho. Instead of constant vowel qualities, the acoustic data indicate that there are substantial differences in vowel quality, in particular, for three of the four vowels F_1 is appreciably higher in tense syllables, indicating an auditorily more open vowel. It is interesting to note that the formant difference between lax /u/ and tense /ṵ/ is substantially less after the labial consonant /p/ than after /t/. Figure 4 makes the vowel quality differences more evident by representing their mean positions on a standard formant plot, as in Ladefoged (1982). Although the vowels in the tense syllables are placed lower than the lax ones on this chart, we should note that it does not necessarily follow that they have lower tongue positions. These acoustic differences are also compatible with the tense vowels having a shorter vocal tract length due to the larynx having

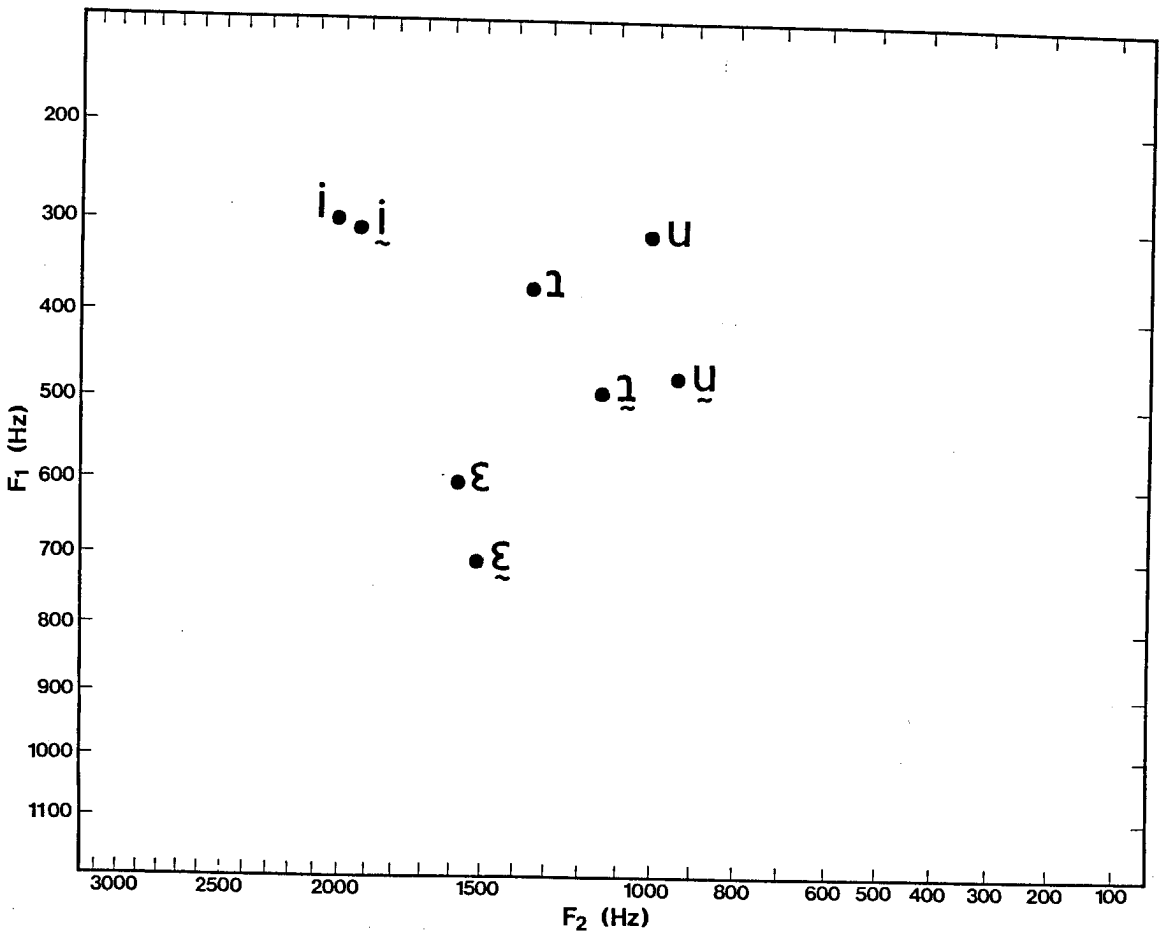


Figure 4. A formant chart showing the difference between the Hani lax and tense vowels in the words in Table 4. The points represent a mean of two tokens for each vowel, except for /ɛ/ and /ɛ̃/ for which we have only single tokens.

been raised. Note that the direction of the differences is the reverse of that associated with the tense/lax difference in Germanic.

In addition to the formant frequency differences, vowels in tense syllables are shorter than in lax ones (mean duration of the lax ones is 539 msec, of tense ones 421 msec), and tense syllables may terminate in very strong glottalization or a glottal stop. F_0 also shows differences. In falling tone syllables the onset to the pitch contour is generally higher in tense syllables than in lax ones, and the pitch does not fall to such a low level, perhaps because of the shorter vowel. In mid tone syllables, the pitch is slightly higher for tense vowels but the difference is small. We did not find any consistent differences in the consonants in the two sets of words.

The aerodynamic data also indicate that there is the the same kind of difference in phonation type as in Jingpho. The airflow was measured in 3 readings of the wordlist in Table 4. In lax vowels the mean flow was 112 ml/sec, and in tense vowels it was 98 ml/sec, a difference that is highly significant ($p < .001$), despite the fact that one pair, the second pair in Table 4, may have been consistently misread on these occasions since it shows the reverse of the pattern seen with all the others. We do not have any pressure data for Hani, so we cannot be sure that the increased flow in the lax vowels was not due to an increase in overall respiratory effort. But it seems unlikely that there should be an increased respiratory effort in lax vowels; and the acoustic data provide no support for such a supposition (the lax vowels do not sound louder). Nor can we explain the increased flow in lax vowels as being due to changes in the supralaryngeal tract shape. As we have noted, the lax vowels have lower first formants. If these are due to supralaryngeal differences, they must imply closer articulations, which would decrease rather than increase the flow rate in lax vowels. It seems quite clear that the difference in flow rate is more likely to be due to a difference in phonation type than to any supralaryngeal difference. This difference in phonation type may also be associated with a raised larynx in the tense vowels, which, as we have noted, is compatible with the acoustic evidence.

(c) Yi

Yi is the name in current use for several of the minority peoples of the mountainous regions of Southwestern China. The name used by the speakers of the best known of the languages spoken by these peoples is Nasu (or variants). Nasu is a cluster of dialects in the Loloish group (Thurgood 1982). A substantial amount of data was obtained from one speaker, the linguist Zhang Ting Xian, who comes from Luquanqiu in the Northern part of Yunnan Province, China. Supplementary data were obtained from two other speakers from the same area; for one of these speakers only aerodynamic data is available. We will refer to this data as Luquan Nasu, or, more simply, as Yi. Luquan Nasu has three tones similar to those of Jingpho and Hani, but in this case there are 14 vowels arranged in seven tense-lax pairs, each of which can occur on all three tones. Illustrative words are given in Table 5.

Table 5
Words illustrating tense and lax in Luquan Nasu (Yi).

lax		tense	
pi 55	"shut"	p̰i 55	"small chicken"
pʰe 55	"house"	pʰḛ 55	"cheat"
vi 33	"far"	v̰i 33	"blossom"
bo 31	"bright"	bo̰ 31	"depend on"
lɯ 31	"beg"	l̰ɯ31	"shake"

The distinction between tense and lax syllables in Yi is different in some respects from either of the other languages discussed above; in particular, it affects the initial consonants of the syllable more. As exemplified in the spectrograms of the eight pairs of words in Figure 5, aspirated stops have a longer period of aspiration in tense syllables. This is contrary to what was observed in Jingpho. In addition, voiced stops, including prenasalized ones, have longer voicing in tense syllables. This pattern is similar to that observed for the stress differences of English, where voiceless stops are more aspirated and voiced stops more voiced in stressed syllables than in unstressed ones (Lisker and Abramson 1967). In Figure 5, note that the formant frequencies of the /a/ vowels are in much the same position in both sets, indicating that there are no differences in vowel quality or variations in the shape of the vocal tract due to pharyngealization or larynx raising in these examples.

In passing, we should also comment on the prenasalized breathy voiced stops illustrated in this figure. The voiced aspiration in these stops is unlike that in Indo-Aryan languages such as Hindi which are more breathy and less voiced, or in Southern Bantu languages such as Tsonga which are fully voiced and have less audible friction. In Yi, there is regular voicing during the nasal and the very brief stop closure. The release of the stop is accompanied by voicing vibrations, but also by a large amount of (presumably) glottal friction.

No overall differences in vowel quality seem to occur with the tense and lax phonations, although one speaker had a markedly lower vowel for /ḛ/ than for /e/. No duration differences between tense and lax vowels were observed in high and mid tone syllables, but in falling tone syllables tense vowels were significantly shorter than lax vowels, perhaps because falling pitch combines with tense phonation to produce an earlier complete glottal occlusion. As for F_o, there is a slightly higher mean pitch in tense syllables than in lax ones. High tone syllables are frequently somewhat falling in pitch; there is a greater fall in pitch for tense syllables with high tone than for lax high-toned syllables, largely since they start from a higher level. (These findings do not agree well with those of Dantsuji (1982) on the Xi-de dialect of Yi. He reported that the tense vowels are lower in quality, as in Hani, and have lower F_o.)

The aerodynamic data indicate that there are a number of differences among our three speakers of Yi, particularly in relation to the airflow in the words

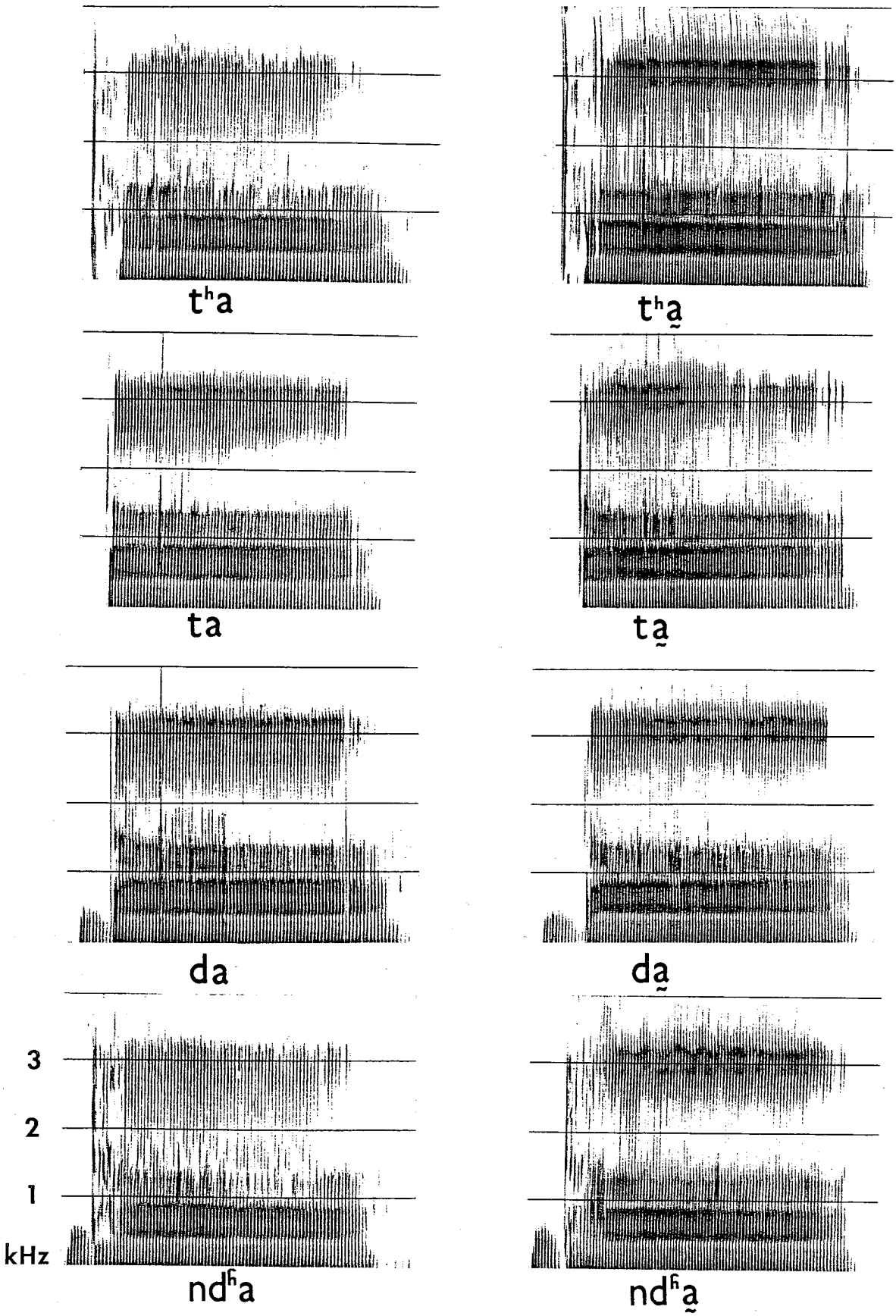


Figure 5. Wide band spectrograms illustrating the differences between lax and tense syllables in Yi. All these items were spoken on mid tone.

examined. Zhang Ting Xian usually had a greater flow rate during the lax vowels than during the tense vowels; in 16 pairs of words the mean difference is 14.2 ml/sec. This is despite the fact that the recorded pressure (which was only available during the bilabials) was much lower (on the order of 30 mm H_2O lower) in lax than in the tense syllables. We are therefore sure that for him there is a difference in phonation type, the tense vowels being pronounced with a greater glottal constriction. A second speaker had no reliable differences in the airflow in the two sets of words. The third had no difference when recorded on one day but on the second day produced the tense vowels with greater airflow than the lax vowels. Unfortunately we have no pressure records accompanying the airflow records for the second and third speakers, but the acoustic records from the session with the third speaker in which the tense vowels had greater airflow show that these vowels also have greater intensity, and were presumably produced with a greater respiratory effort on this occasion. If this were a relatively normal occurrence it would be consistent with the similarity between the VOT patterns in Yi tense syllables and English stressed syllables pointed out above.

The acoustic records for this third speaker also show that, despite the greater airflow, the glottis was vibrating in a way that produced relatively more energy in the second harmonic (in comparison with the fundamental) in the tense vowels, as was the case in the Jingpho tense vowels discussed earlier. Spectral measures on between 3 and 5 repetitions of 7 pairs of words were obtained; the means for a total of 28 tokens show that the fundamental is 6.2 dB higher than the second harmonic in lax vowels but 6.8 dB lower in tense vowels. This difference is highly significant ($p < .001$). The pair of words $p^he/p^h\tilde{e}$ included in the means, is exceptional. For this speaker, tense /ẽ/ has a higher F_1 than lax /e/. The result in these particular high tone words is that the second harmonic is close to the first formant resonance in the lax tokens, but is below the formant frequency in the tense tokens, and the formant resonance boosts the amplitude of the third harmonic instead. The consequence is a reversal of the expected relationship between fundamental and second harmonic amplitudes in three of the four readings of this pair of words.

Zhang Ting Xian read a different list of words; 19 pairs of words from this list were measured. For this speaker there is almost invariably more energy in the second harmonic than in the fundamental, but there is a smaller mean difference in lax vowels than in tense vowels, 3.4 dB vs 5.9 dB (a net difference of 2.5 dB). Overall, the difference is highly significant ($p < .001$), but on closer inspection an interaction between tone and the lax/tense contrast can be observed. If high tone pairs are examined separately there is no significant difference between the lax and tense vowels (only 0.6 dB net difference, compared with 4.1 dB net difference in mid and falling tone pairs only). This pattern was not found for any other speakers in this study.

(d) Wa

The fourth language examined, Wa, is not a Tibeto-Burman language but belongs to the Palaungic branch of the Mon-Khmer family (Diffloth 1980). Data was recorded from only one speaker, who came from the Lin Tsang district of Yunnan Province, China, again from a village a few miles from the Burmese border. The variety of Wa spoken in this area is usually referred to by the name Kawa. There are eighteen monophthongs, nine of which appear in lax and nine in tense syllables (Dai 1958, cited in Diffloth 1980). The tense/lax contrast occurs in both open syllables and closed syllables, including those closed by /h/ and /ʔ/, as illustrated in Table 6. The contrast does not occur after aspirated consonants which only occur in loanwords. There are no tones in Kawa.

Table 6
Words illustrating lax and tense in Wa.

lax		tense	
pi	"forget"	p̰i	"harmonica"
pe?	"goat"	p̰e?	"you (pl)"
tɛ	"peach"	t̰ɛ	"sweet"
tɛh	"change"	t̰ɛh	"reduce"
pɔ	"don't leave"	p̰ɔ	"psoas muscle"
pu	"thick"	p̰u	"fly"

The acoustic measurements on 11 pairs of words including those in Table 6 reveal no salient distinctions in the formant frequencies of the vowels of the two sets. But measurements on 9 of these pairs (measureable spectra could not be obtained for two pairs) show a consistent difference in the distribution of spectral energy. The fundamental was 1.8 dB higher than the second harmonic in the lax vowels, but 2.0 dB lower in the tense vowels. The difference is highly significant ($p < .005$). Thus, as in the other languages, the acoustic records suggest that there is a difference in phonation type between the two sets of vowels. Measurements of duration showed a general trend for the tense vowels to be a little shorter than their lax counterparts. (In Wa, vowels in syllables checked by /?/ are as long as vowels in open syllables, but vowels before /h/ are considerably shorter than these.) In these citation forms the pitch was uniformly falling in both syllable types; there were no consistent F_0 differences between the lax and tense syllables, other than a weak trend for the contour of tense syllables to start and finish a little lower than that of lax ones. Our data tend to suggest that there are some differences in the consonants in lax and tense syllables; in lax syllables VOT is longer for stops and nasals are longer.

The aerodynamic measures for Wa are somewhat limited, owing to the speaker having some difficulty in saying the words while talking into a face mask (a problem that was not encountered with any of the other subjects). We did however, obtain completely reliable pressure and airflow records for 9 pairs of words. As is evident from Figure 1, which shows two of these pairs of words, the pressure varied somewhat, seemingly in accord with extraneous factors, such as the sequence in which the words were read. As a result, there is no significant difference between the pressures for the lax and the tense vowel sets (the means are 107 and 109 mm H_2O respectively). However, there was a higher mean flow for the lax vowels, 117 ml/sec, than for the tense vowels, 107 ml/sec; a difference that is probably significant ($p < .05$). In general, we can conclude that the lax vowels have a greater rate of airflow for a given pressure, and accordingly must have been produced with a less constricted glottis.

Discussion

The detailed descriptions given above show that the phonetic exponents of the distinction between lax and tense syllables are somewhat different in each of these languages. A summary of the findings is given in Table 7; the order of the languages in this table is arranged to fit in with the discussion which follows.

Table 7
The use of various phonetic possibilities for realizing the differences between lax and tense syllables in the four languages.

	Hani	Yi	Jingpho	Wa
flow/pressure ratio	lax greater	lax probably greater	lax greater	lax greater
ratio of F_o to second harmonic	lax greater	lax greater	lax greater	lax greater
height of F_1	lax lower	no difference apart from E/E for one speaker	no difference	no difference
vowel duration	lax longer	lax longer in falling tone	no difference	lax slightly longer
overall F_o	lax slightly lower	lax slightly lower	no difference	lax slightly higher
F_o onset	lax sometimes rising	no consistent difference	lax rising with high tone	no difference
voice onset time	no difference	lax somewhat shorter	lax longer	lax longer
other consonantal properties	tense: final glottalization (or glottal stop with falling tone)	lax: voiced stops less prevoicing		lax: nasals longer

It may be seen that the only thing in common across all four languages is the use of phonation type, as is shown by the flow/pressure ratio, and the relative energy levels of the fundamental and the second harmonic. The distinction between lax and tense is redundantly signaled by a larger number of properties in some of the languages, particularly the two Lolo-Burmese languages Hani and Yi, with fewer involved in Jingpho and Wa. There is no uniform pattern of association between phonation type and other properties as suggested by Matisoff (1973) and others. We will discuss briefly the significance of our findings from the historical and descriptive viewpoints.

The difference between lax and tense has different origins in the languages concerned. In Hani and Yi, tense vowels are the normal reflex of original final stops (Wheatley 1982, Bradley 1979). On the other hand, the distinction between lax and tense in Wa derives from an original voicing distinction in the consonants preceding the affected vowels. Lax vowels occur in syllables with originally voiced initials, tense vowels in those with originally voiceless initials (Diffloth 1980: 27-28). There are no published studies on the origin of the distinction in Yunnanese Jingpho, but this is evidently another language in which it evolved from initial rather than final consonants. The Yunnanese dialect has lost the voiced series of stops found in Kachin Jingpho. Scott DeLancey (personal communication) confirms that the items we recorded as examples of lax syllables are shown as having voiced initials in Hanson's dictionary of Kachin (Hanson 1906). Graham Thurgood (personal communication) agrees with our interpretation on the basis of examination of our wordlist and comparison with a recently published Jingpho-Chinese dictionary.

The languages we have studied thus divide into two pairs. Two of them, Hani and Yi, have tense vowels derived from former checked syllables, and two of them, Jingpho and Wa, have lax vowels derived from former voiced initials. Some of the phonetic differences in the realization of the contrast between lax and tense in the languages may result from this fact. In those Tibeto-Burman languages which retain the smooth/checked distinction, such as Burmese (Javkin and Maddieson 1983), Lahu (Hombert 1983), and Jingpho (see above), the vowels in the checked syllables are substantially shorter. In our data, vowel duration is a prominent factor in the tense/lax difference in the two languages which derive tense vowels from checked syllables. It is absent or negligible in the other two languages. In our data we notice also that Hani and Yi are the two languages in which we find tendencies for an overall lower pitch and for auditorily higher vowels (lower F_1) to occur in lax syllables. Again, these phenomena can be matched in related languages which have retained the checked/smooth contrast. For example, checked syllables in Burmese have the highest onset pitch level (Javkin and Maddieson 1983: 116) and noticeably different vowel allophones from other syllables; lower for high and mid vowels, higher for low vowels. This pattern can be seen in the spectrograms in Figure 6 from 2 speakers of Burmese. This pattern of shorter and (usually) lower vowels in closed syllables is widespread (Maddieson 1985), and is parallel to the development of vowels in closed syllables in languages like English and German, except that in this case the resulting vowels have been traditionally referred to as "lax". In the Tibeto-Burman case there is the additional factor of the phonation tension to consider, and hence these vowels have been called "tense". The mechanism by which such allophones originate is, however, likely to have something in common in the two cases, and may be associated with the phenomenon of target undershoot as described by Lindblom (1963). Figure 6 suggests how this might occur; note how most of the duration of the very short vowel in /sî?/ is affected by the transition from the initial consonant, but the longer vowel in /sî/ has a large portion beyond the influence of the initial consonant.

When the contrast between lax and tense results from loss of an initial voicing contrast we would anticipate somewhat different though partly parallel results. Most obviously, some distinction in the pitch patterns, especially at the onset of the vowel might be expected. This is found in Jingpho, and, in particular, with high tone syllables - exactly where the effect of voicing differences has been shown to be greatest in tone languages (Hombert, Ohala and Ewan 1979: 41; Maddieson 1984). Neither Jingpho nor Wa show an overall lower F_0 in lax syllables, although this effect is found in some other Mon-Khmer languages

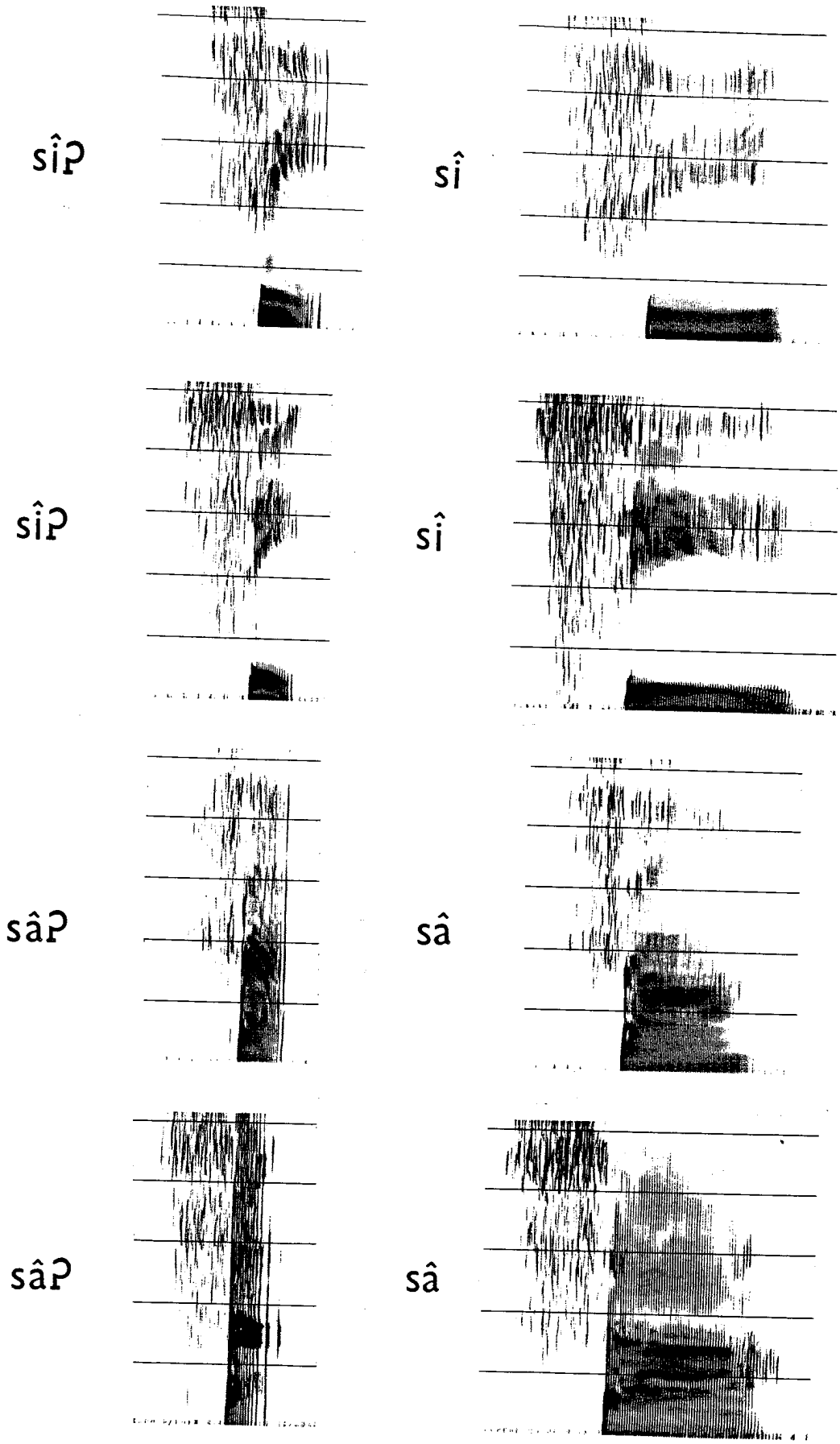


Figure 6. Wide band spectrograms of four Burmese words spoken by one male and one female speaker illustrating vowel duration and quality differences in open vs. glottal-stop-final syllables.

e.g. Mon (Lee 1983). As for vowel length, the lack of a difference in Jingpho and Wa is unsurprising since it is a contrast in the preceding consonants that is concerned. Although no difference in vowel quality was observed in Jingpho or Wa in our data, it might be noted that the most natural expectation is that the former voiced initials would lead to higher rather than lower vowels in the lax category as a result of the pharyngeal cavity expansion which often accompanies voicing in consonants (Westbury 1983, Ewan 1979). This result is approximately what is found in Brou (Miller 1967). It is, however, different from that which is found in Mon, where, to the extent that any vowel quality difference is found, it is divergent for different vowels (Lee 1983).

Thus the varying patterns for the realization of the contrast between lax and tense in the four languages we have studied can be understood in historical terms, at least in part. But how should these patterns be described in a synchronic phonology? In each case there are grounds for believing that the prime distinction is that of phonation type and that other characteristics are subordinate. The other properties are usually contingent on the occurrence of particular tones or consonants or vowels and hence should be regarded as allophonic rules. These rules must obviously differ from language to language. For example, a rule that shortens tense vowels in falling tone syllables is required for Yi but not in any of the other languages. Jingpho requires a rule deriving a rising onset to a high tone in a lax syllable. Only Hani has higher F_1 for tense vowels. And so on. In other words, although the tense/lax contrast may be signalled by a complex of features in each language, there is no uniformity across languages in the way that this is done. In short, there is no support for tense and lax as general phonological terms.

Although this is not apparent from the aerodynamic or spectral measures, we also have some evidence that the phonation contrast itself should be described differently in Hani and Yi as opposed to Jingpho and Wa. Note that in Jingpho and Wa there is a longer voice onset time for stops in lax syllables. Delayed onset of voicing can arise from several causes; in this case it seems most likely to occur because the vocal cords are more widely spread in lax syllables than tense ones. It is certainly unlikely that it is due to greater vocal cord tension in the lax syllables in these languages. On the other hand, in Yi there is a longer VOT for the tense syllables, and in Hani no difference in VOT. This should be considered together with the tendency for higher overall pitch in the tense syllables in this pair of languages. Tensing the vocal cords raises pitch and can result in delayed voice onset. This tension could also result in a raised position for the larynx, producing the lower F_1 for tense vowels in Hani and, sometimes, in Yi. A possible interpretation of these differences is that the phonation contrast in Jingpho and Wa is between a relatively more breathy phonation (lax) and a "normal", or modal, type of phonation (tense), whereas in Hani and Yi the contrast is between a relatively more laryngealized phonation (tense) and a modal phonation (lax). Modal phonation is thus assigned a different role in the phonologies of these languages in a way that corresponds to the different historical origins of the distinction between lax and tense syllables. These considerations suggest that, rather than using a feature [tense/lax], the contrast may be described by adapting laryngeal features proposed by Halle and Stevens (1971). In Jingpho and Wa vowels are [+ slack] (lax) or [- slack] (tense), whereas in Hani and Yi vowels are [- stiff] (lax) or [+ stiff] (tense).

Conclusion

As we have shown, even languages in a particular linguistic area may differ substantially in the way that they arrange the clusters of phonetic properties that signify the contrast between their lax and tense vowels. And the tense/lax contrast can involve quite unrelated phonetic properties when languages from different areas are considered. While the terms "lax" and "tense" may sometimes be a useful shorthand in a linguistic description, it is necessary to spell out exactly what is to be understood by them in each case.

Acknowledgments

We gratefully acknowledge the assistance of many linguists in the People's Republic of China, and of Graham Thurgood and Scott DeLancey in the USA. The research reported in this paper was supported by a grant from the National Institute of Neurological and Communicative Disorders and Stroke to the UCLA Phonetics Laboratory and by funds provided by the Academic Senate of the University of California, Los Angeles.

References

Anonymous. 1959. [An Outline of Jingpho Grammar] (In Chinese). Minority Language Research Bureau, Chinese Academy of Sciences, Peking.

Bell, Alexander Melville. 1867. Visible Speech: The Science of Universal Alphabetics. London.

Bickley, Corine. 1982. Acoustic analysis and perception of breathy vowels. Working Papers, Speech Communication Group, MIT 1: 71-81.

Bradley, David. 1979. Proto-Loloish (Scandinavian Institute of Asian Studies Monograph Series 39). Curzon Press, London and Malmö.

Catford, J.C. 1977. Fundamental Issues in Phonetics. Indiana University Press, Bloomington.

Chao, Yuan-Ren. 1930. A system of tone letters. Le Maître Phonétique 30: 24-27.

Chomsky, Noam and Morris Halle. 1968. The Sound Pattern of English. Prentice-Hall, New York.

Dai Qinxia. 1958. [On tense and lax vowels] (In Chinese) Shao shu min zu yu wen lun ji. Shanghai.

Dantsuji, Masatake. 1982. An acoustic study on glottalized vowels in the Yi (Lolo) language - a preliminary report. Studia Phonologica 16: 1-11.

Diffloth, Gérard. 1980. The Wa Languages (Linguistics of the Tibeto-Burman Area 5.2). California State University, Fresno.

Disner, Sandra F. 1983. Vowel Quality: The Relation Between Universal and Language-Specific Factors (UCLA Working Papers in Phonetics 58.) University of California, Los Angeles.

Egerod, Søren. 1971. Phonation types in Chinese and Southeast Asian languages. Acta Linguistica Hafniensa 13: 159-171.

Elert, Claes-Christian. 1964. Phonologic Studies of Quantity in Swedish. Skriptor, Uppsala.

Ewan, William G. 1979. Laryngeal Behavior in Speech (Report of the Phonology Laboratory 3). University of California, Berkeley.

Fant, Gunnar. 1983. Preliminaries to analysis of the human voice source. Quarterly Progress Report 4: 1-27. Speech Transmission Laboratory, Royal Institute of Technology, Stockholm.

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

Gregerson, Kenneth J. 1976. Tongue root and register in Mon-Khmer. In Philip N. Jenner et al (eds) Austroasiatic Studies, Volume 1. University of Hawaii Press, Honolulu.

Halle, Morris and K. N. Stevens. 1971. A note on laryngeal features. Quarterly Research Report, Research Laboratory of Electronics (M.I.T.) 101: 198-213.

Hanson, O. 1906. A Dictionary of the Kachin Language. Rangoon.

Hansson, Inga-Lill. 1984. A comparison of Akha, Hani, Khàtú, and Pîjò. Paper presented at 17th International Conference on Sino-Tibetan Languages and Linguistics, University of Oregon, Eugene.

Henderson, Eugenie. 1967. Grammar and tone in South East Asian languages. Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig: Gesellschafts- und Sprachwissenschaftliche Reihe 1/2: 171-178.

Hombert, Jean-Marie. 1983. A brief encounter with Lahu tones. Linguistics of the Tibeto-Burman Area 7.2: 109-111.

Hombert, Jean-Marie, John J. Ohala and William G. Ewan. 1979. Phonetic explanations for the development of tones. Language 55: 37-58.

Hu Tan and Dai Qinxia. 1964. [Tense and lax vowels in Hani] (In Chinese) Zhongguo Yuwen 128: 76-87.

Jakobson, Roman, Gunnar Fant and Morris Halle. 1952. Preliminaries to Speech Analysis. Acoustics Laboratory, Massachusetts Institute of Technology, Cambridge, Mass.

Javkin, Hector and Ian Maddieson. 1983. An inverse filtering analysis of Burmese creaky voice. UCLA Working Papers in Phonetics 57: 115-125.

Jespersen, Otto. 1889. The Articulation of Sounds, Represented by Means of Analphabetic Symbols. Elwert, Marburg.

Jones, Daniel. 1956. An Outline of English Phonetics. Heffer, Cambridge.

Jørgensen, H-P. 1969. Die gespannten und ungespannten Vokale in der norddeutschen Hochsprache, mit einer spezifischen Untersuching der Struktur ihrer Formantenfrequenzen. Phonetica 19: 217-245.

Kirk, Paul, Peter Ladefoged and Jenny Ladefoged. 1984. Using a spectrograph for measures of phonation types in a natural language. UCLA Working Papers in Phonetics 59: 102-113.

Ladefoged, Peter. 1964. A Phonetic Study of West African Languages. Cambridge University Press, Cambridge.

Ladefoged, Peter. 1967. Three Areas of Experimental Phonetics. Oxford University Press, London.

Ladefoged, Peter. 1981. The relative nature of voice quality. Journal of the Acoustical Society of America 69, Supplement 1: S67.

Ladefoged, Peter. 1982. A Course in Phonetics (2nd. ed.). Harcourt Brace Jovanovitch, New York.

Ladefoged, Peter. 1983. The linguistic use of different phonation types. In Diane Bless and James Abbs (eds) Vocal Fold Physiology; Contemporary Research and Clinical Issues: 351-360. College Hill Press, San Diego.

Ladefoged, Peter and Norma Antoñanzas. 1984. Computer measurements of breathy voice quality. Journal of the Acoustical Society of America 75: S8.

Ladefoged, Peter and Anthony Traill. 1984. Instrumental phonetic fieldwork. In Jo-Ann Higgs and Robin Thelwall (eds) Topics in Linguistic Phonetics in Honor of E.T. Uldall (Occasional Papers in Linguistics and Language Learning 9): 1-22. The New University of Ulster, Coleraine.

Lee, Thomas. 1983. An acoustical study of the register distinction in Mon. UCLA Working Papers in Phonetics 57: 79-96.

Li Yungsuei 1979. [survey of Hani] (In Chinese) Minzu Yuwen (Peking) 2: 134-151.

Lindau, Mona. 1978. Vowel features. Language 54: 451-563.

Lindblom, Björn. 1963. Spectrographic study of vowel reduction. Journal of the Acoustical Society of America 35: 1773-1781.

Linell, Per, Björn Svensson and Sven Öhman. 1971. Ljudstruktur. Gleerups, Lund.

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

Maddieson, Ian. 1984. The effects on F_0 of a voicing distinction in sonorants and their implications for a theory of tonogenesis. Journal of Phonetics 12: 9-15.

Maddieson, Ian. 1985. Phonetic cues to syllabification. In Victoria A. Fromkin (ed) Phonetic Linguistics. Academic Press, New York.

Maran, La Raw. 1971. Burmese and Jingpho: A Study of Tonal Linguistic Processes (Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics 4). University of Illinois, Urbana.

Matisoff, James A. 1973. Tonogenesis in Southeast Asia. In Larry M. Hyman (ed) Consonant Types and Tone (Southern California Occasional Papers in Linguistics 1): 71-96. University of Southern California, Los Angeles.

Miller, J. D. 1967. An acoustical study of Brou vowels. Phonetica 17: 149-177.

Perkell, Joseph S. 1969. Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study. M.I.T. Press, Cambridge, Mass.

Perkell, Joseph S. 1971. Physiology of speech production: a preliminary report on two suggested revisions of the features specifying vowels. Quarterly Progress Report, Research Laboratory of Electronics (M.I.T.) 102: 123-139.

Pols, Louis C. W. 1977. Spectral Analysis and Identification of Dutch Vowels in Monosyllabic Words. Institute for Perception, Soesterberg.

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

Schachter, Paul and Victoria Fromkin. 1968. A Phonology of Akan: Akuapem, Asante, Fante (UCLA Working Papers in Phonetics 9). University of California, Los Angeles.

Spa, J. J. 1970. Generatieve fonologie. Levende Talen 266: 191-204.

Stevens, K. N., Jay Keyser and Haruko Kawasaki. 1984. Toward a phonetic and phonological theory of redundant features. In J. S. Perkell and D. H. Klatt (eds) Symposium on Invariance and Variability in Speech Processes. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Stockwell, Robert P. 1973. Problems in the interpretation of the Great English vowel shift. In M. E. Smith (ed) Studies in Linguistics in Honor of George L. Trager. Mouton, The Hague: 344-362.

Sweet, Henry. 1877. A Handbook of Phonetics. Clarendon Press, Oxford.

Sweet, Henry. 1908. The Sounds of English. Clarendon Press, Oxford.

Thurgood, Graham. 1980. Consonants, phonation types, and pitch height. Computational Analyses of Asian and African Languages 13: 207-219.

Thurgood, Graham. 1982. Subgrouping on the basis of shared phonological innovations: a Lolo-Burmese case study. Proceedings of the Eighth Annual Meeting of the Berkeley Linguistics Society, University of California, Berkeley: 251-260.

Wängler, Hans-Heinrich. 1961. Atlas Deutscher Sprachlaute (2nd. ed.). Akademie-Verlag, Berlin.

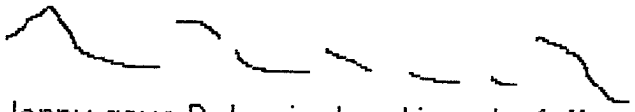
Westbury, John. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. Journal of the Acoustical Society of America 73: 1322-1336.

Wheatley, Julian K. 1982. Comments on the 'Hani' dialects of Loloish. Linguistics of the Tibeto-Burman Area 7.1: 1-38.

Macintosh usage for linguists

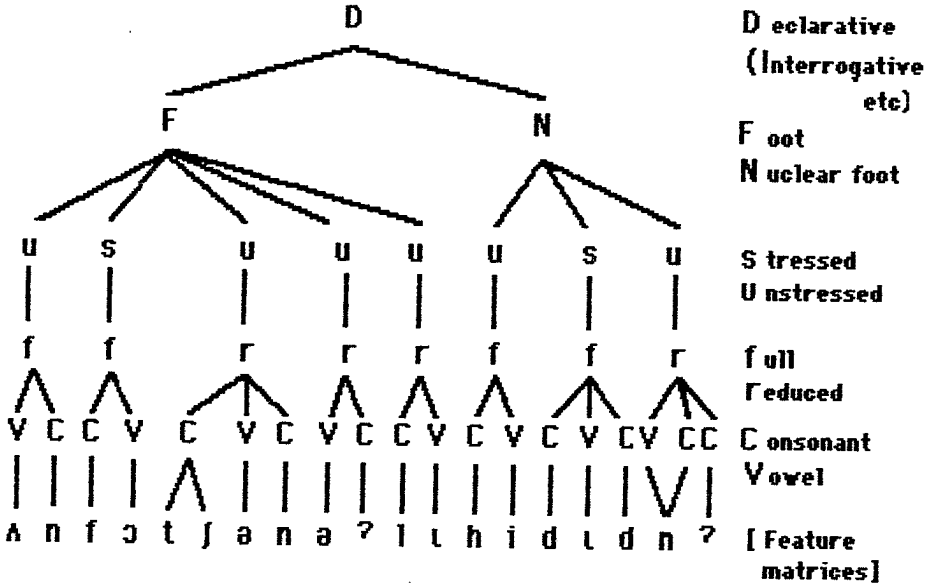
Peter Ladefoged

The whole of this page has been written on a Macintosh computer, using the regular Apple Imagewriter as a printing device. The Macintosh is clearly a very suitable computer for linguists. Macwrite is a reasonable word processor (though certainly not the best in the world); and, taken together with Macpaint, it allows one to do all sorts of fancy things, such as giving accurate representations of intonation curves:



(1) Jenny gave Peter instructions to follow

or drawing trees showing new forms of CV or metrical phonology:



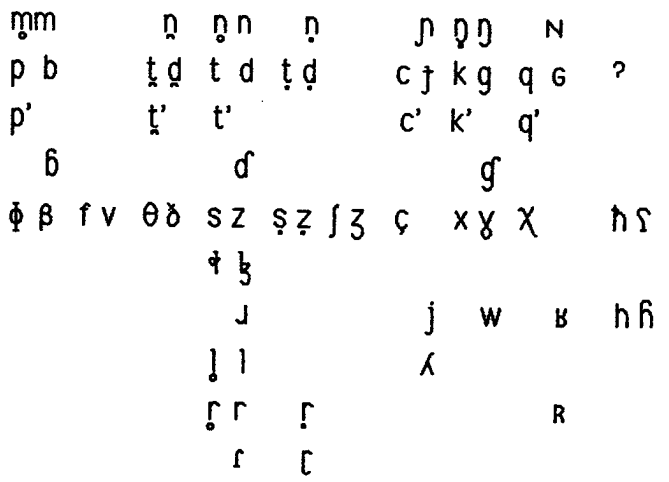
(2) " Unfortunately he didn't "

(There is not really anything particularly new about the above tree. It just makes evident the relationships inherent in a traditional view of English phonology, by stating the tiers on which choices are available. Other forms of CV phonology that consider stress to be a multi-valued feature, and conflate the [Stress-Unstressed] and the [full-reduced (vowel)] tiers are, in some senses, only notationally different. But, as noted by Vanderslice and Ladefoged (1972), multi-valued representations of stress distort the phonetic facts and lead to lost or improper generalizations.)

Of course, the techniques for drawing phonological relations can also be used for language family or syntactic trees, if you happen to like doing that sort of thing.

As the above illustrations show, phonetic symbols are available for the Macintosh computer. The Macintosh system includes a wide range of fonts, all of them rather curiously named after cities. This part of the text is in 12 point Geneva. There are two phonetic fonts available from Megatherium Enterprises (P.O. Box 7000-417, Redondo Beach, CA 90277), called LGeneva and LNew York, after the corresponding system fonts. The consonant chart below is in a modifed form of LGeneva which, following the Macintosh tradition, I have called Edinburgh. (I could not call it London, because that is the name for an existing font.)

Consonant chart



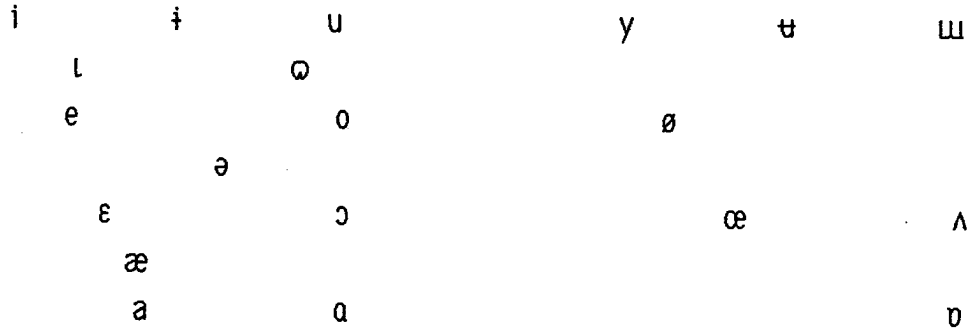
Symbols in the Edinburgh font, but not in the LGeneva font from which it is derived include:
ʐ̥ ɻ̥ l̥ ɽ̥ ɮ̥ ɴ̥ ŋ̥ ɽ̥ ɓ

In addition, the symbols ɸ ɣ ʒ ð in the Edinburgh font are ɸ ɣ ʒ ð in LGeneva, and all the descenders in the Edinburgh fonthave been lengthened to form g ɠ j ɟ p q ɥ ʃ instead of g ɠ j ɟ p q ɥ ʃ.
Additional symbols in both fonts include:
š ž č ǰ ř ʇ ʗ ɭ̥ m̥ n̥ r̥ ɬ (in LGeneva ɬ),
and a number of symbols used by linguists in special areas such as:
p̊ t̊ c̊ k̊ q̊ ɤ ɕ þ

The principal weakness of the Macintosh system is that it is not possible to superimpose diacritics on a symbol; there is no way of doing the equivalent of backspacing and overtyping. As a result every symbol such as a vowel with an accent or tone mark above it, or a consonant with a diacritic indicating voicelessness below it, must be in the font as a separate item. This disadvantage is somewhat mitigated by the fact that each font such as LGeneva or LNew York has 218 printing characters; and

each of them exists in six different sizes, which can be used as superscripts or subscripts as appropriate. All the common vowels in the charts below also have various modified forms such as: a á à â ä ã ă ā ą ą́ ą̀.

Vowel charts



The symbol for a rhotacized vowel ɚ is also available. The Edinburgh font contains both ʊ and ʊ ; only ʊ is available in LGeneva.

As with all Macintosh fonts, one can use these fonts not only in different sizes, but also in different styles. If one has a mind to, one can write transcriptions

ɪn ɪ'tælɪks **ɔɹ ɪn 'boʊld 'feɪs** ɔɹ 'ʌndəlaɪnd

'aʊtlaɪnd ɔ 'ʃædoʊd or in any combination, 'lɑdʒ ɔ 'smɔl

If one has a font editor it is fairly simple to make one's own symbols, as I have done for the Edinburgh font. This, of course, prompts the question as to why one should buy the two fonts LGeneva and LNew York, which together form the **Mac the linguist** software available from Megatherium Enterprises. But I am sure it is well worth the modest expenditure ($50), as there is an enormous amount of labor involved in making 218 symbols in each of two fonts in each of six sizes. If I had not had them (and a font editor program), I would never have been able to produce the Edinburgh font which I prefer. It is also very convenient to have full documentation, listing every symbol and its relation to the keys on the keyboard, each of which has four possibilities, the key by itself, with the shift key depressed, with the option key depressed, and with both the shift and the option keys depressed. **Mac the linguist** should have a great future.

Reference

Vanderslice, R. and P. Ladefoged. 1972. Binary Suprasegmental features. **Language 48.4,819-838.**

Digital Inverse Filtering for Linguistic Research

Hector Raul Javkin, Norma Antonanzas-Barroso and Ian Maddieson

1. Introduction

Speech waveforms are the product of both the phonatory setting and the shape of the vocal tract. In a voiced sound, the wave shape produced by the vibrating glottis is modified by the particular shape of the vocal tract above the glottis. The setting of the glottis can be varied to produce different glottal wave shapes, and many languages make significant use of different modes of glottal vibration (Ladefoged 1983). It is not easy to make direct observations of what the larynx is doing during speech or to observe directly how the pattern of air flow through the glottis differs from utterance to utterance or from language to language. However, if the effect of the vocal tract (basically, the formant pattern) can be subtracted from the speech waveform, we can examine the glottal waveform itself without requiring any invasive procedure. The technique which performs this analysis is known as inverse filtering. Inverse filtering can be performed using a specially designed analog filter (Miller 1959) or can be implemented digitally using a computer.

In this paper we will describe the digital inverse filtering methods used at the UCLA Phonetics Laboratory for the recovery of glottal air volume velocity from speech waveforms. Our research goals are to learn both how the glottal waveform contributes to linguistic contrasts and what differences there are between different speakers making the same linguistic contrast. For this reason, our system was designed to enable a relatively large number of languages and speakers to be conveniently studied. In order to reach conclusions about what is typical of a particular language, it is necessary to study a group of speakers of that language rather than a single individual. The same methodology could also be used for studying laryngeal function in clinical populations.

2. Recording and sampling

In order to recover the glottal pulse shape accurately the original speech signal has to be recorded and sampled without phase distortion throughout the frequencies of interest. Because standard tape recorders distort phase, the recording must be made with an FM system or by direct digital conversion. Standard studio microphones also introduce considerable phase distortion, so the signal has to be captured either with an instrumentation condenser microphone or with a suitable airflow mask. Each of these systems has its own advantages and disadvantages. Instrumentation condenser microphones have the necessary phase characteristics and a frequency response which permits the recording of considerable detail in the glottal pulse, but they cannot capture the DC component of the airflow. This means that the microphone will miss the continuous part of the airflow that occurs when the glottis vibrates without making a complete closure during the cycle. In addition, the amplitude of glottal flow cannot be calibrated when using a microphone. An airflow measurement mask can preserve the DC component and give a calibrated, quantitative measure of flow. But the relatively low frequency response of the airflow mask, means that it cannot capture the high-frequency details obtainable with the microphone. These two recording methods can therefore be seen as complementary, which is why we have used both. There is a difference between them in the filtering process that is used. With a microphone, the effect of radiation from the lips must be taken into account. The mask eliminates this effect and hence removes one step from the inverse filtering process.

2.1 Using a microphone

An instrumentation condenser microphone (Bruel and Kjaer 4133) is used with an appropriate cathode follower and amplifier. The low-frequency response of this microphone requires special care in recording. As mentioned above, a microphone signal has to undergo integration in order to remove the effects of lip radiation. Integration produces a low frequency emphasis with a slope of 6 dB per octave. As a result, any low frequency noise in the recording will be enhanced greatly. Low frequency noise is extremely difficult to eliminate from a recording environment. Our recordings are made in a sound treated booth inside a room cleared of all extraneous noise. Furthermore, because of a low frequency resonance found in our building's ventilation ducting, all building ventilation is turned off. During recording, subjects hold the microphone approximately one centimeter to the side of one edge of the lips. This position avoids any direct impact of airflow on the microphone. The short distance makes it possible to reduce amplifier gain, further reducing noise.

2.2 Using an airflow measurement mask.

Recording with an airflow measurement mask requires fewer precautions than using a microphone. The airflow measurement system constructed by Martin Rothenberg has a useful frequency response from zero to about 1000 Hz (Rothenberg 1973). Use of the mask automatically eliminates the effect of radiation at the lips, so that the signal need not (and must not) be integrated in the inverse filtering analysis. As a result, low frequency noise presents a much less serious problem, and recordings can be made in any relatively quiet environment.

2.3 Recording

We use two different FM recording methods: a Tandberg FM recorder with a frequency response from zero to about 5000 Hz, and a locally constructed FM encoding and decoding circuit which permits recording with a standard, high quality tape recorder. This latter system has a frequency response from zero to about 2000 Hz. The tape recorder has to have an excellent tape transport system (excellent "wow-and-flutter" characteristics), since variations in frequency caused by variations in tape speed will cause amplitude distortion in the decoded signal. With both kinds of recordings, the signal is continuously monitored on an oscilloscope for the possibility of clipping or excessively low amplitude.

2.4 Passing the signal into the computer

Computer sampling for inverse filtering requires an anti-aliasing filter with minimal phase distortion. Our sampling method is designed for a sampling rate of 18 kHz, the rate of the slower of our two analog-to-digital converters. This sampling rate means that we can only deal with frequencies below 9 kHz in the signal if we are to avoid aliasing (according to the Nyquist theorem). Therefore the signal must be low-pass filtered. Bessel design filters have excellent phase characteristics but relatively gradual rolloff. In order to attenuate the signal sufficiently at 9 kHz, the filter chosen must have a low cutoff frequency. We installed a 3 kHz 6-pole Bessel low-pass filter for all signals intended for inverse filtering, so that the signal is attenuated 25 dB at the Nyquist rate. A 16-bit analog-to-digital converter passes the signal into the computer with a far higher signal-to-noise ratio than the rest of the recording and data input steps.

3. Inverse filtering program

As outlined in the introduction, the output of the vocal folds is altered by passing through the vocal tract and, except when a mask is used, by passing out of the vocal tract into the air surrounding the speaker (Fant 1970). To recover the glottal waveform it is necessary to remove these effects. A schematic view of this process is shown in Figure 1 (based in part on Figure 1 in Fant 1983b). The rise and fall of airflow reflected in the recovered glottal waveform is closely related to the changes in glottal area occurring during the vibratory cycle (Fant 1983a). The glottal wave we are interested in recovering is, of course, a function of time, but the formants introduced by the vocal tract, as well as the effect of lip radiation, are best understood in terms of frequencies and bandwidths. We therefore developed the inverse filter in the frequency domain with the objective of applying it in the time domain. The z-transform allows us to pass between these two domains.

3.1 The z-transform

The definition of the z-transform, when applied to a finite sequence, is (Oppenheim & Schafer 1975):

$$X(z) = \sum_{n=0}^{n=N-1} x(n) * z^{-n}$$

where

- n = the index of the sequence
- x = the amplitude of the signal at sample point n
- z = a complex variable which is a function of frequency
- N = length of the sequence

Two fundamental properties of the z-transform, derived from its definition, greatly simplify the conversion from one domain to the other. To discuss these we will let Z represent the operation of z-transformation. The properties are: (a) Linearity. The transformation of a sum is equivalent to the sum of the transformed elements. That is:

$$Z(x(n)+y(n)) = Z(x(n)) + Z(y(n))$$

And, (b), the shifting property, by which a shift in the time domain can be expressed in the frequency domain as the transform multiplied by z raised to the shift. That is:

$$Z(x(n-k)) = z^{-k} * Z(x(n)) = z^{-k} * X(z)$$

This property has some very convenient results. It means that equations of filters developed in the frequency domain as powers of z can be converted to shifts in time, or delays. A delay of one in the sample point means taking the immediately preceding sampled value.

3.2 The vocal tract model

The inverse filtering program reverses the supra-glottal effects in the speech signal. To remove the formants and the effects of radiation at the lips, it is necessary to construct a filter which has the inverse response. Following

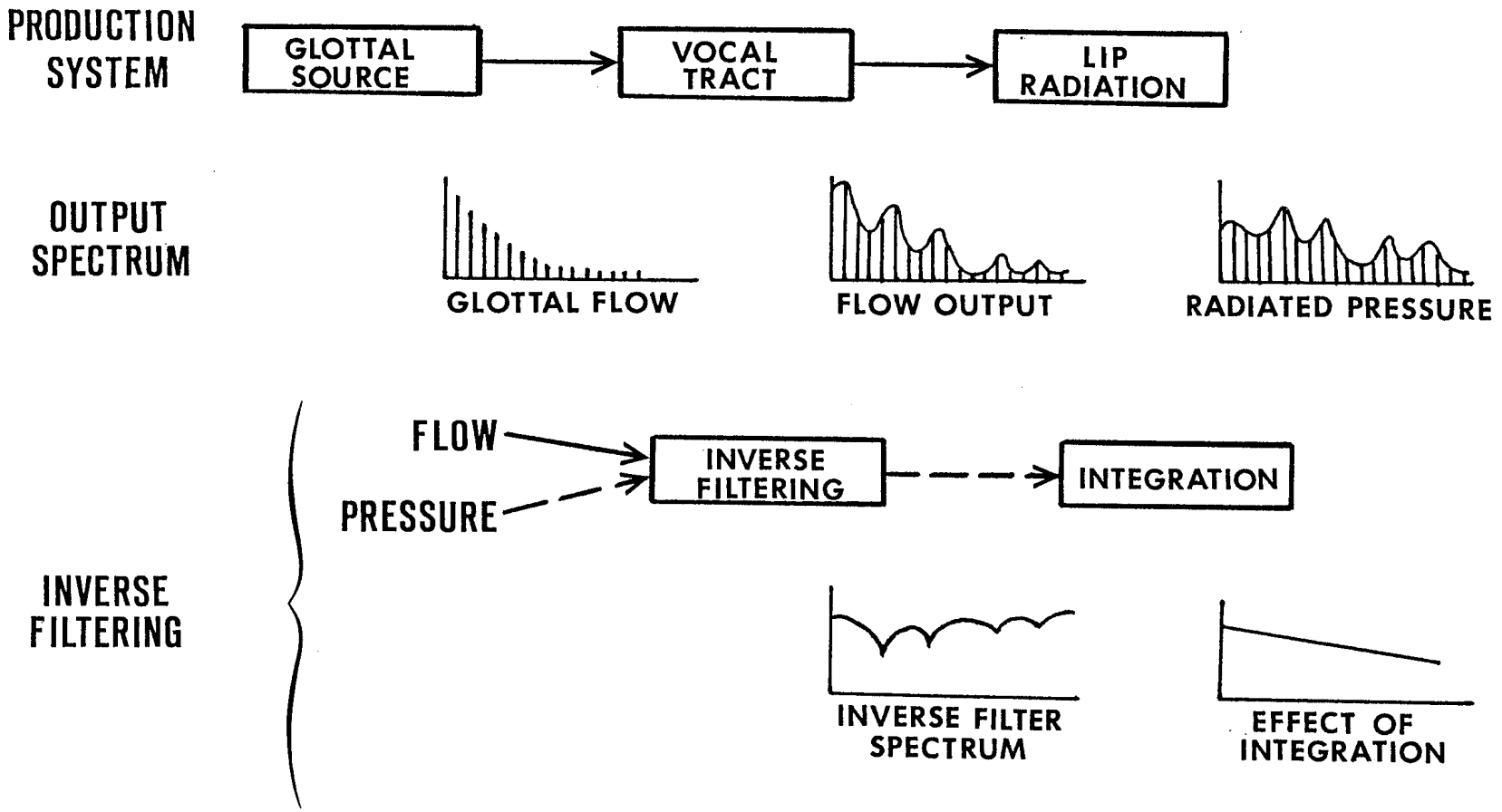


Figure 1. Outline of the speech production system and the output spectrum at different stages (above), and the steps in inverse filtering a pressure wave (dotted arrow) or a flow wave (solid arrow) together with their effects in the frequency domain.

Fant (1970) and Rabiner and Schafer (1978), the digital filter which models the vocal tract is:

$$VT(z) = \prod_{k=1}^{k=M} \frac{1 - e^{-b_k T} * 2\cos(f_k T) + e^{-2b_k T}}{1 - e^{-b_k T} * 2\cos(f_k T) * z^{-1} + e^{-2b_k T} * z^{-2}}$$

where

M = the number of formants
b_k = the (one-sided) bandwidth in radians of formant F_k
f_k = the formant frequency in radians of formant F_k
T = the sampling period

This means that the effect of the vocal tract can be modeled by a cascade of second-order systems, each representing a formant.

$$F(z) = \frac{1 - e^{-bT} * 2\cos(fT) + e^{-2bT}}{1 - e^{-bT} * 2\cos(fT) * z^{-1} + e^{-2bT} * z^{-2}}$$

To invert a formant, the numerator and denominator are simply reversed, yielding:

$$IF(z) = \frac{1 - e^{-bT} * 2\cos(fT) * z^{-1} + e^{-2bT} * z^{-2}}{1 - e^{-bT} * 2\cos(fT) + e^{-2bT}}$$

(IF = inverted formant)

To invert the effect of the vocal tract, we must invert the effect of all the formants.

If there is a finite number of inverted formants, the frequency response of the system will have a high frequency emphasis. Each inverted formant will raise the response of frequencies higher than its characteristic frequency, as can be seen in Figure 2 below:

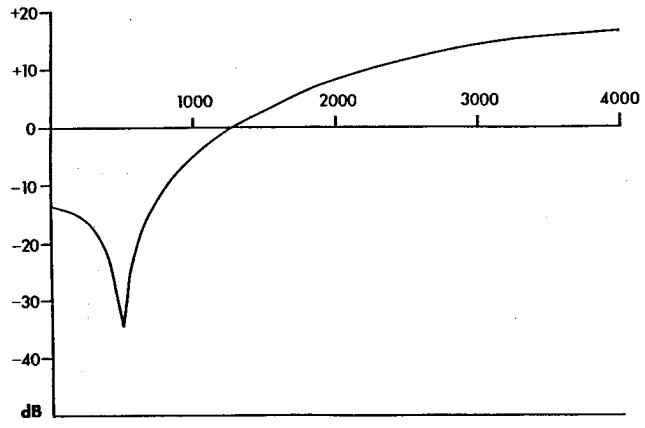


Figure 2. Effect of inverted formant in the frequency domain.

Fant (1959, 1970) solved this problem for analog filters by adding a "higher pole correction" in cascade with the rest of the system. Gold and Rabiner (1968) found a different solution in the response characteristics of digital filters, which are symmetrical around the Nyquist rate and periodic at the sampling rate. This means that an inverted filter with formants up to the Nyquist rate is effectively a filter with an infinite number of formants. Each of the higher formants flattens the response of those with lower frequencies, yielding a correct filter response. Using a sample rate of 18 kHz, which means a Nyquist rate of 9 kHz, this calls for nine inverse formants, from 500 to 8500 Hz. Since the higher formants are included only for the purpose of maintaining a correct response, only the four lowest formants are varied to match the speech signal. Methods for finding values for these four formants are described in section 4. Formants 5 through 9 have much less energy in the speech signal, and are found by the formula:

$$F_k = \frac{(2k-1)500}{L/17.5}$$

where

F_k = frequency of a particular formant
k = number of the formant
L = assumed length of the vocal tract in centimeters

3.3 The effect of the lips

In addition to the effects of the vocal tract, the glottal signal also undergoes the effects of radiation at the lips, unless the mask is used. Following Markel and Gray (1976), this radiation can be modelled as:

$$L(z) = 1 - z^{-1}$$

Inverting the effect of radiation at the lips is somewhat less straightforward than inverting the effect of the vocal tract. Radiation at the lips amounts to a differentiation. Although integration inverts the effects of differentiation, there are some constraints on how the integration can be performed. Simply inverting the formula for lip radiation would yield:

$$IL(z) = \frac{1}{1 - z^{-1}}$$

At zero frequency, z will equal 1, and the value would be infinity. This means that low frequencies will be greatly amplified, provoking an unstable response. Multiplying z by a constant (k) which is less than 1 will prevent the value from becoming infinity. The expression thus becomes:

$$IL(z) = \frac{1}{1 - kz^{-1}}$$

Multiplying the expression by 1 - k will make the amplification (or gain) equal to 1 at zero frequency. This is desirable because zero frequency represents the DC component of air flow, which is transmitted with a gain of 1 by the vocal apparatus. The expression that will invert lip radiation properly is therefore:

$$IL(z) = \frac{1 - k}{1 - kz^{-1}}$$

Inverting the effects of both the lip radiation (when appropriate) and the vocal tract will yield the glottal pulse.

3.4 Implementation

To apply the inverse filter to a time-varying waveform, the equations given above have to be transformed into the time domain. An input X(z) is processed by the system function IF(z) to produce an output Y(z), as represented below:

X(z) --->| IF(z) |----> Y(z)

This can be expressed as:

$$Y(z) = IF(z)X(z)$$

So that:

$$IF(z) = Y(z)/X(z)$$

where

X(z) = z-transform of the input
IF(z) = the filter equation in z (the frequency domain)
Y(z) = z-transform of the output

Recall the equation for inverting a single digital formant:

$$IF(z) = \frac{1 - e^{-bT} * 2\cos(fT) * z^{-1} + e^{-2bT} * z^{-2}}{1 - e^{-bT} * 2\cos(fT) + e^{-2bT}}$$

We can simplify the expression with the following substitutions.

$$A = -e^{-bT} * 2\cos(fT)$$

$$B = e^{-2bT}$$

$$D = 1 - e^{-bT} * 2\cos(fT) + e^{-2bT}$$

This yields:

$$IF(z) = \frac{1 + A * z^{-1} + B * z^{-2}}{D}$$

Substituting the quotient for IF(z) we get:

$$\frac{Y(z)}{X(z)} = \frac{1 + A * z^{-1} + B * z^{-2}}{D}$$

Distribution of the denominator and multiplying both sides by X(z) and distributing again, yields:

$$Y(z) = (1/D) * X(z) + (A/D) * z^{-1} * X(z) + (B/D) * z^{-2} * X(z)$$

Recall the properties of the z-transform described in 3.1. The linear property allows us to transform the parts of the sum individually. The shifting property means that a power of z in the frequency domain becomes a delay in the time domain. The transformation yields:

$$y(n) = (1/D) * x(n) + (A/D) * x(n-1) + (B/D) * x(n-2)$$

Transforming the lip radiation equation into the time domain is similar. Let us recall the lip radiation function:

$$IL(z) = \frac{1 - k}{1 - kz^{-1}}$$

Substituting for IL(z) yields:

$$\frac{Y(z)}{X(z)} = \frac{1 - k}{1 - kz^{-1}}$$

This yields:

$$Y(z) * (1 - kz^{-1}) = (1 - k) * X(z)$$

Once again, the z-transform changes powers of z in the frequency domain into delays. This yields, after distribution:

$$y(n) - k * y(n-1) = (1 - k) * x(n)$$

And the final expression, as used in the computer program, is:

$$y(n) = k * y(n-1) + (1 - k) * x(n)$$

The program implements the inversion of the 9 formants in cascade. The signal passes through them in descending order. Although the order is not crucial, we preferred to process the lowest formants last so that these would suffer the least computer round-off error.

The program processes the input file in 10 msec (180 point) segments. To handle formant changes during speech the filter settings for the four lowest formants are set for each segment. The program uses a special start up procedure at the beginning of the file and at the beginning of each segment. Each filter in the cascade requires 2 successive points before it can properly filter the third. The first filter processes 18 points. Only points 3 through 18 are filtered correctly. The first filter therefore passes only points 3 through 18 to the

second filter. The second filter uses points 3 and 4 for its own delay and passes points 5 through 18 to the next filter. Each succeeding filter thus uses properly filtered input, discards the first two points it receives and passes through only properly filtered output. After point 18, the entire cascade is producing properly filtered output until the segment of 180 points is filtered.

In order to prevent discontinuities in the output caused by changes in filter settings between one segment and the next we use overlapping segments. Each succeeding segment goes through the same start-up procedure as the first. The 18 points taken from the input file by the second segment for starting up overlap the last points processed by the first segment. This is repeated with each suceeding segment until the entire file is processed.

4. Analysis procedures

In order to analyze a sufficient number of speakers per language, and a sufficient number of words per speaker, as well as an interesting number of languages, the time required for analysis has to be reduced as much as possible without sacrificing accuracy. Setting formants and bandwidths for inverse filtering and analyzing the output have been automated as much as possible.

4.1 Formant tracking.

The need for a fast and accurate method led to the development of a two step process. LPC analysis and finding the roots of the LPC polynomial provides a rapid way of obtaining formant frequencies. However, the bandwidths obtained by the LPC procedure reflect several factors including losses in the vocal tract, radiation at the lips (in a pressure wave), and the rolloff in the glottal source spectrum. This rolloff is generally estimated as being on the order of 12 dB per octave for modal voice, but it is variable with particular glottal settings. Since the objective of our inverse filtering analysis is to study the glottal source, this contribution to formant bandwidths should be discounted in inverse filtering. Thus we aim to supply formant bandwidths for the inverse filter which reflect only a monotonic increase in formant bandwidth with increasing formant number. We are still unsure of a principled way to calculate these bandwidths appropriately, and have sometimes used several values and selected the best-looking filtered output post hoc. Also LPC analysis does not invariably find formants at all and only the appropriate locations. The LPC analysis therefore provides only initial estimates of formants. These estimates, provided at 10 msec intervals for an entire input file, are written in a format accessible to a text editor, and then reviewed and edited for errors. The edited file is then used by the inverse filtering program to calculate filter settings for every 10 msec window of data.

4.2 Measurement of the inverse filtered waveform.

As mentioned in the introduction, recovering the glottal signal as a waveform is not in itself the ultimate aim of performing inverse filtering. Rather, the goal is to obtain a series of measures which characterize the signal and can be used to distinguish different modes of phonation. We have developed a program that yields many such measures, once a user has marked two points on each glottal pulse. For measurement, the pulse is defined as consisting of a closure portion (possible of zero duration), rising portion and a falling portion, as can be seen in the idealized pulse in Figure 3.

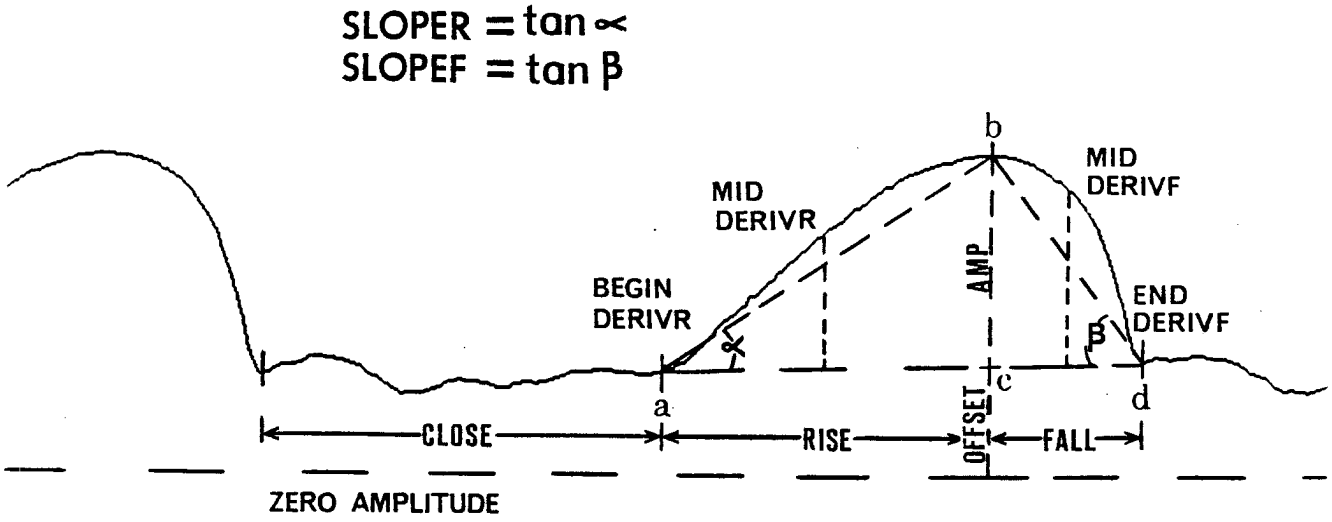


Figure 3. Sample recovered glottal flow waveform, marked for measurement.

On a display of the filtered waveform, The user marks the beginning of the rising portion (A) and the end of the falling portion (D) of each pulse. The program then produces a number of measures for each glottal pulse:

TIME:	Starting time of the pulse, defined as the beginning of the closed portion.
CLOSED:	Duration of the closed portion
RISE:	Duration of the rising portion
FALL:	Duration of the falling portion
OFFSET:	DC offset of the baseline of the pulse from zero.
AMP:	Relative maximum amplitude of the pulse: i.e.: maximum amplitude minus the offset
MID DERIVR:	Middle Rising Derivative: the average derivative along 7 points centered in the middle of the rising portion.
MID DERIVF:	Middle Falling Derivative: the average derivative along 7 points centered in the middle of the falling portion.
BEGIN DERIVR:	Maximum Derivative during the first millisecond of the rising portion of the pulse.
END DERIVF:	Maximum Derivative during the last millisecond of the falling portion of the pulse.
SLOPER:	Slope of a line from A to B in the diagram above.
SLOPEF:	Slope of a line from B to D in the diagram above.

Note that the program provides three types of measures of the slopes of the rising and falling portions. The mid derivatives represent a mean rate of airflow increase and decrease in the pulse, whereas the slopes characterize a triangular approximation to the pulse shape. The begining and end derivatives characterize the sharpness of the pulse onset and offset.

A number of additional measures may be calculated from the measures above. The sum of the duration of the rising and falling portions provide the duration of the open portion. Ratios which express the relationship of different portions of each pulse to each other can also be calculated, such as the ratio of open to closed duration and ratios of the rising and falling slope measures. The sum of

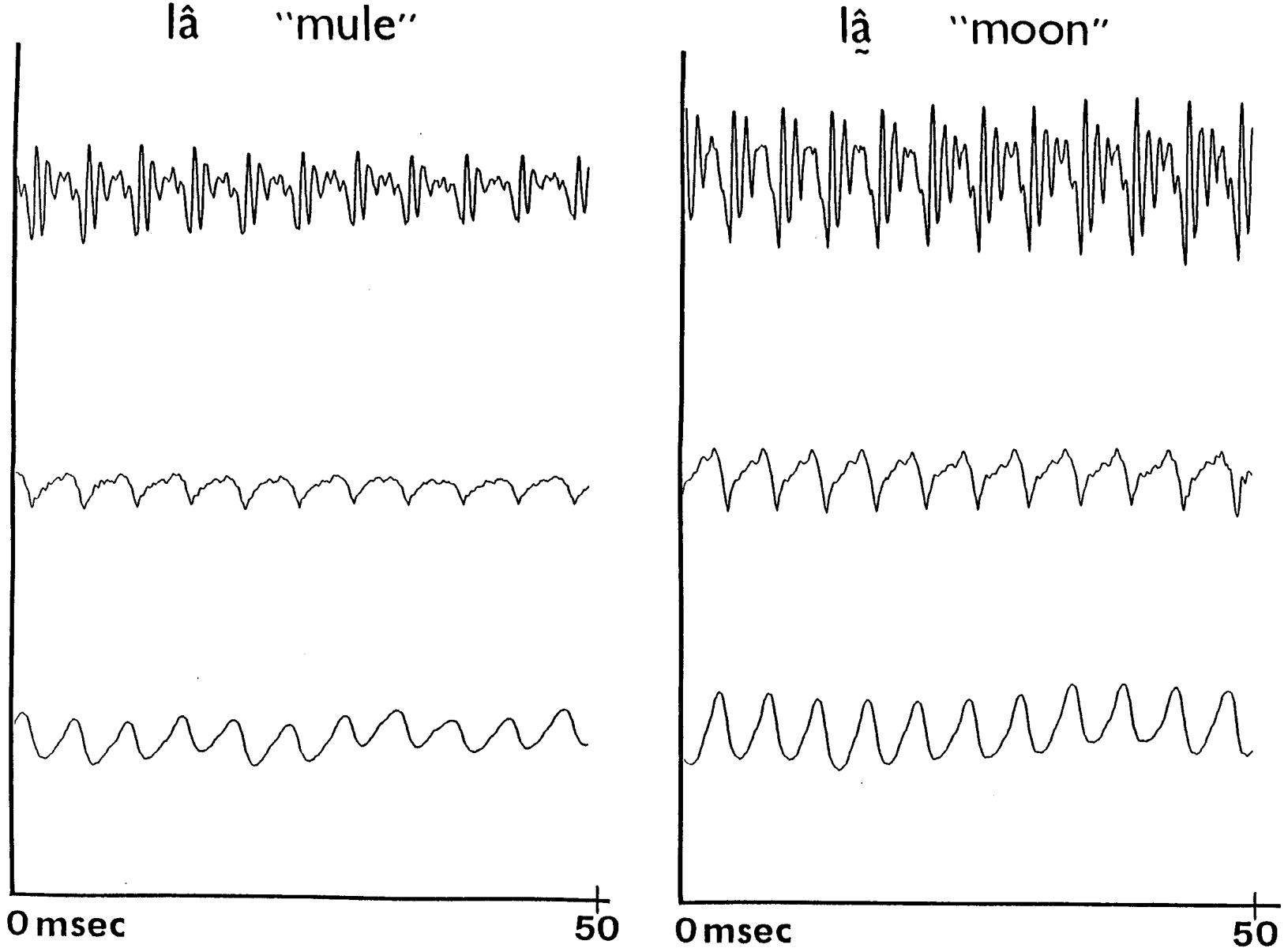


Figure 4. Sample pressure waveform, inverse filtered waveform and integrated inverse filtered waveform of "smooth" and "creaky" tone words in Burmese.

closed, rising and falling portions provide the pitch period, from which the fundamental frequency is calculated. Similarly, measures of jitter (the variation of the duration of one period compared to the next) and shimmer (period to period variations in airflow amplitude) can be computed.

5. A sample analysis

As an illustration of the methods we have used, consider the sample analysis shown in Figure 4. This figure shows 50 msec excerpts from two words of Burmese spoken by a female speaker and recorded using the microphone and FM recorder. The words differ in that the one on the right (/la/ "mule") has the "smooth" tone, whereas the one on the left (/la/ "moon") has the "creaky" tone. These words have similar pitch contours, but the smooth tone word is longer whereas the creaky tone word is pronounced with an increasingly tense larynx. The sample from the smooth tone word is taken from somewhat later in the word than the creaky tone sample, hence it has somewhat lower amplitude. These analyses were prepared using a completely algorithmic procedure. Values for formants were obtained using an LPC technique; relevant portions of this analysis are reproduced in Table 1.

Table 1. Formant values used in sample analysis

	Time	F1	F2	F3	F4
/la/	432	915	1234	2999	4399
"mule"	442	933	1320	2957	4383
	452	924	1394	2968	4352
	462	906	1397	2984	4348
	472	913	1437	3034	4346
/la/	352	937	1641	2979	4458
"moon"	362	954	1553	2953	4336
	372	977	1450	2912	4317
	382	990	1430	2920	4299
	392	978	1334	2869	4311

The table shows quite large differences between successive formant values; for example, note the falling F2 in "moon". As a result, the inverse filter for each window of data is quite different. Bandwidths of k * 50 Hz (where k = the number of the formant) were supplied for these formants, and the inverse filtering routine was run.

In the figure, the top line shows the digitized speech waveform. The second line shows the inverse filtered waveform before integration, and the third line is the integrated signal. The nominal time origin of the three waveforms in each set is the same, but since the filtering processes build in small delays, the displayed waveforms have a small offset in time from each other. With more active editing of the formant values a smoother output might be achieved, with fewer ripples in the filtered waveforms - but it is not easy to decide whether such ripples are due to lack of smoothness in the glottal movements or to a residue left from imperfect filtering.

The 10 complete glottal periods in each of the samples in the figure were marked and measured using the measurement program described in section 4.3 above. Although the slopes are quite different between these two samples, this difference is not important, since it is a consequence of the difference in amplitude. The ratio of the rising and falling slopes in the two samples is the same. However, there is a significant difference in the duration of the closed

portion of the pulse. In the smooth tone sample the closed portion has a mean duration of about 9% of the open portion of the cycle, whereas in the "creaky" sample the closed portion has a mean duration of about 16% of the closed portion.

6. Conclusion

The methods described in the preceding pages can provide detailed and accurate data on the output of the vocal folds. The automatic procedures we have developed make the process manageable and make it possible to examine the productions of a larger number of speakers. With accurate data on a sufficient number of speakers, meaningful conclusions about the linguistic uses of differences in vocal fold vibration can be drawn.

Acknowledgments

We owe a great deal, in this effort, to Gunnar Fant, as well as to Peter Ladefoged, Marie Huffman and other members of the Phonetics Laboratory at UCLA. We have benefited greatly from discussions with Corine Bickley, Franklin Cooper, Richard Hamming, Steven Hunt, Michael O'Malley, Lloyd Rice and Hisashi Wakita.

References

Cooley, J.W., P.A.W. Lewis & P.D. Welch (1969) The finite Fourier transform. IEEE Transactions on Audio and Electroacoustics. Vol. AU-17.2: 77-85

Gold B. & L. Rabiner (1968) Analysis of digital and analog formant synthesizers. IEEE Transactions on Audio and Electroacoustics. Vol AU-16: 81-94.

Fant, G. The acoustics of speech (1959) Proceedings of the Third International Congress on Acoustics: 188-201. Reprinted in: R.W. Schafer & J.D. Markel (eds.) Speech Analysis. New York: IEEE Press 1978.

Fant, G. (1970) The Acoustic Theory of Speech Production. The Hague: Mouton.

Fant, G. (1983a) Preliminaries to analysis of the human voice source. Quarterly Progress and Status Report (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm) 4/1982: 1-27.

Fant, G. (1983b) The voice source - acoustic modeling. Quarterly Progress and Status Report (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm) 4/1982: 28-48.

Hamming R.W. (1977) Digital filters. Englewood Cliffs: Prentice Hall.

Ladefoged, Peter (1983) The linguistic use of different phonation types. In D. Bliss & J. Abbs (eds) Vocal Fold Physiology; Contemporary Research and Clinical Issues: 351-360. San Diego: College Hill Press.

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

Markel, J.D. & A.H. Gray (1976) Linear prediction of speech. New York: Springer-Verlag.

Oppenheim A.V. & R.W. Schafer (1975) Digital signal processing. Englewood Cliffs: Prentice Hall.

Rabiner L.P. & R.W. Schafer (1978) Digital processing of speech signals. Englewood Cliffs: Prentice Hall.

Rothenberg, M.A. (1973) A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. Journal of the Acoustical Society of America 53.6:1632-1645.

Schafer R. & J. Markel (1979) Speech analysis. New York: IEEE Press.

Redefining the scope of phonology

Peter Ladefoged

Abstract

The functions of speech and language interact in a way that has a considerable effect on our view of phonology. If we are mainly concerned with the way language functions to convey objective information we will pay particular attention to the phonological oppositions that distinguish meaningful units such as words and phrases. But if we are more concerned with the sociolinguistics and attitudinal information conveyed by the sounds, we will have to pay attention to phonetic details that are not used for indicating phonological oppositions. The implications of these differences are discussed, and it is suggested that a viable phonology that is concerned only with strictly linguistic patterning will be somewhat different from the current view of phonology.

Speech has many functions; and not all of them are part of language. Similarly language has many functions; and not all of them are part of speech. One of the major functions of language that is not part of speech is to act as a mirror for the world. Language provides us with symbols for our experiences, and thus gives us ways of grouping these experiences into categories, and ways of qualifying and relating one experience to another. As a result we are able to form concepts and manipulate ideas. Some kinds of thinking may be possible without language, but we could never develop scientific theories without words. Our language acts as a model of the world as we know it, much as a map serves as a model for a piece of the terrain, enabling us to plan journeys. This function of language has little or no effect on our views of phonology - though it does, obviously, have a considerable effect on how we view semantics.

Of much greater importance for our view of phonology is the fact that one of the functions of both speech and language is the conveying of objective information (or misinformation). In order to convey this information, meaningful units have to be distinguished from one another. Of course, when we consider language as a model of our world, the symbols used for categorizing and relating our experiences have to be distinguished from one another as well. But when language is functioning in this way it does not matter how the distinctions are achieved. Language acts as our mirror of the nature of things, irrespective of whether the words are written, spoken, or simply mental images. We do not need to speak the words of our thoughts.

In order to convey objective information in spoken form, languages typically contrast between 20 and 35 segmental phonemes (Maddieson 1984), supplemented by a few suprasegmental devices. This is a comparatively small number out of the total set of possibilities available. On the basis of listening tests I have previously estimated (Ladefoged 1967) that it is possible to distinguish about 50 vowels in the plane of the primary cardinal vowels, i.e. in which front vowels are unrounded and back vowels are rounded, with the degree of rounding being predicted from the height. Adding the possibility of front rounded vowels and back unrounded vowels, together with so-called under-rounded and over-rounded vowels (such as the Assamese low back vowel which has the close lip rounding normally associated with a vowel such as [u]) would almost double this number. With the addition of nasalized vowels (which, even with training and experience, are not as distinct as oral vowels), rhotacized (r-colored) vowels, and other

possible secondary articulations, the total number of distinguishable voiced monophthongs (to add further constraints) is undoubtedly well above 100. This is all without considering various types of diphthongs that are traditionally considered as single segments with on-glides or off-glides. We must also note possible variations in phonation type. Many languages distinguish sets of laxly voiced vowels from regularly voiced vowels (e.g. Jingpho and other languages of Southeast Asia). Others (such as Mpi) contrast laryngealized and non-laryngealized vowels. In calculating the total number of vowels we should consider each vowel as potentially occurring on three different phonation types; it is not at all difficult to distinguish at least this number of different voice qualities.

When we consider all these possibilites it seems that Shaw (1920) may have considerably underestimated the number of distinguishable vowels. In the play Pygmalion (and in the My Fair Lady version, Lerner and Loewe, 1956) Colonel Pickering expresses admiration for the expert phonetician, Henry Higgins, who is able to distinguish 130 different vowels, as opposed to Pickering's 24. Shaw probably based his estimate of the number of vowels on his reading of Sweet (1890), the acknowledged prototype for Higgins, who provided 72 distinct symbols for vowels without considering diphthongs, differences in phonation types, and other aspects of vowel quality for which he provided diacritics. I would estimate that an expert phonetician could distinguish more than 250 vowels of all types. And what an expert phonetician can distinguish, anyone who has been brought up speaking a language that uses one of these distinctions can do just as well.

The number of distinct consonants is also considerable, even if we limit ourselves fairly strictly to what must be called single segments (i.e. disregarding all affricates, prenasalized stops, etc, although many of them function as single phonological segments). The IPA (1979) chart has 81 symbols for consonants, without taking account of oppositions such as that between dental and alveolar stops (which contrast in many Australasian languages), voiceless nasals (as in Burmese), or differences between aspirated and unaspirated obstruents (Sindhi has 25 stop consonants, only 10 of which appear as distinct symbols on the IPA chart). We must also consider all the secondary articulations, such as labialization, palatalization, velarization and pharyngealization, which would far more than double the number of possibilites. And again we have to note differences in phonation type, as well as airstream mechanisms of the kind that form clicks, ejectives and implosives. A very conservative estimate would place the total number of consonantal segments as being up in the hundreds, making the total number of possibly contrasting segments as high as 600-800. A comparable number occurs in Maddieson's (1984) survey of the phonological segments that occur in 317 languages carefully selected so as to exemplify the range of the world's languages. He found that when he considered each phonological segment to be represented by its principal allophone he had to recognize about 650 phonetically distinct segments, without considering variations in length.

There are many reasons why languages do not use such a large number of segmental oppositions. Perhaps the most important is that they are not necessary; languages can have a sufficiently large stock of morphemes while using only a small number of segmental oppositions. The phonological devices used by languages do not require the wealth of phonetic possibilities.

Many of the subtle distinctions that can be made among sounds are used by some of the other functions of speech. Whenever we talk, we convey not only

information about the topic under discussion, but also information about the sociolinguistic group to which we belong. Some of this information is conveyed in the same way as distinctions among words, using differences among phonemes. Thus Porter (1936) describes a case of different individuals with the same set of possible phonemic contrasts, using them in different words, as in "You say [iðər] but I say [aiðər]." (It should be noted that Porter's phonetic observations are not always reliable. He correctly observes that some people pronounce the word "tomato" as [tə'meɪtoʊ], whereas others say [tə'mɑtoʊ]. But he further claims to have observed the word "potato" pronounced as [pə'tɑtoʊ]. This seems very unlikely.) A great deal of sociolinguistic information is conveyed in more subtle ways. It is rigidly codified, although not in terms of discrete oppositions of the kind used in phonemic oppositions. It is difficult to say exactly what degree of diphthongization in the vowel in "mate" marks a person as belonging to a particular social class in London (or Australia, or anywhere else that uses this distinguishing characteristic). But anyone familiar with the regional accents in question can easily place a speaker by pronunciations of this kind.

Because this information is codified in speech in an arbitrary way, we may want to regard it as part of language. But there is no reason to expect the sociolinguistic information to be conveyed by the same aspects of speech sounds as those that are used for distinguishing the linguistic oppositions discussed above. Evidence has been accumulating recently that demonstrates quite conclusively that languages and dialects are differentiated from one another in ways that are not used to distinguish oppositions within any single language. For example Ladefoged and Bhaskararao (1983) have demonstrated differences in the retroflex stops in Hindi and those in Telugu that depend on features of speech that are not used to distinguish oppositions within any single language. Similar points with respect to differences among fricatives have been made by Ladefoged and Wu (1984). Cross-linguistic differences in phonation types that characterize different languages have been described by Lindau (1982). These and many other papers suggest that the sociolinguistic functions of speech cannot be described entirely in terms of the same features as those that are used for describing phonological oppositions. There is no way in which small but reliable differences in retroflexion, or fricative noise, or phonation type can be expressed in terms of phonological feature classifications. Nevertheless, this is what standard feature theories attempt to do, becoming continually more complicated as a result. Thus Jakobson and Halle (1956) can express more phonetic detail than Jakobson, Fant and Halle (1952); Chomsky and Halle (1968) add still more features; and Halle and Stevens (1971) add complexities so as to be able to describe phonetic differences between languages. All this is done in order to be able to describe how the speech of one group of people differs systematically from that of another group of people. But there is no theoretical or empirical reason to expect speech systems to use the same devices for phonological and sociolinguistic purposes.

I do not mean to imply that speech systems never use the same devices for linguistic and sociolinguistic purposes. Quite obviously the ways in which vowels are distinguished within a language often involve the same mechanisms as those used for distinguishing the vowels of one accent from another. Thus the feature Vowel Height (or High) may be used phonologically for classifying vowels or, by means of its scalar values, for phonetic descriptions. But many differences between languages are not of this kind. There is no phonological feature system that allows for the degree of phonetic detail necessary for characterizing the differences between the fricatives in English, Pekingese, Tamil, and Polish (Ladefoged and Wu 1984). Attempting to specify phonetic detail in phonation types

and stop consonants led Halle and Stevens (1971) into proposing a feature system that is reasonable for characterizing phonetic differences but is unacceptable for phonological classification (Anderson 1978). The Halle-Stevens proposal replaces the feature Voice by a set of four features, Stiff, Slack, Spread, and Constricted, so that they can characterize the phonetic difference between, for example, English [p] as in "spy" and the Korean so-called lax [p]; but the cost is that they no longer have the more phonologically useful opposition voiced-voiceless.

There are also other aspects of speech that cannot be expressed in terms of any of the traditional sets of phonological features. Another function of speech is to convey the attitude of the speaker to the topic under discussion, to the person addressed, and, indeed, to the world in general. It is not at all clear how much of this is codified. Some parts of the speaker's attitude to the topic under discussion are normally considered as conveyed by systematically different intonations. Thus we speak of statement versus question intonation in different languages, and emphatic and non-emphatic statements and questions. But what about sarcastic or simpering intonations? Are they part of language?

Similarly, it is not clear how we should consider emotional effects. There seems to be something in common to expressions of anger, astonishment, sorrow, doubt, and love, for example, in many different languages. But they are probably not the same in all languages. An angry Frenchman does not sound like an angry German. But is the anger part the same, and the differences due to the regular linguistic differences between the two languages? And how do we separate out the universal tendencies from those that are plainly learned, cultural, aspects of behavior? Many Englishmen consider it normal to speak in a phlegmatic way with a narrow intonation range that Americans consider as indicative of boredom. The Navaho tend (by American English standards) to speak very softly. As Nihalani (1983) has pointed out, Indian English typically sounds rude or aggressive to speakers of British or American English. All these are learned aspects of the culture. But do we want to consider them part of language?

Yet another aspect of speech, which probably nobody would call a function of language, is that it signals the identity of the speaker. When I walk into a house and call out "Hi, it's me" this is all the information I am conveying. I am not really making a declarative statement. Everyone recognizes that the personal information is not part of language. But it is sometimes not clear what should be regarded as personal information and what is codified socio-linguistic information. We each speak in the way that we do partly because our particular vocal organs have certain characteristics, but also because we choose to use, within limits, our own personal style of speech. Often what might seem to be a personal characteristic of a particular speaker is in fact something that he or she has chosen to copy, which is shared by a small sociolinguistic group. Where does the family unit end, and the local group begin?

Summing up so far, the basic question is how much of all these different kinds of information conveyed by speech do we want to consider as part of language? We can get some help on this problem by considering the differences between spoken and written language. Is it appropriate to speak of a language being <u>reduced</u> to writing - this implying that some part of spoken language is not present in the written form? Or would it be better to say that (virtually) all that is <u>language</u> can be expressed in speech or in writing - and all the sociolinguistic, emotional, and personal, information that is left out is not part of what we want to define as language?

Writing conveys some but not much sociolinguistic information. It is impossible to tell from these printed pages whether this paper has been written by an Englishman or an American, or indeed, by a speaker of one or many other forms of English. You could gain some sociolinguistic information if I were to use certain marked phrases or lexical items such as talking about a full stop as opposed to a period. When the written language does convey sociolinguistic information it does so by means of precisely those phonological devices that are used to convey information about the topic under discussion. We do not need anything beyond a feature system that is capable of identifying linguistic oppositions to handle sociolinguistic information of this sort.

Writing also conveys certain aspects of intonation. From the syntax, morphology, word order and punctuation (period, comma, query, quotes, parentheses, italics and space marks) we can determine something (but far from everything), about the intonation that a given sentence could have. As Bolinger (1977) has pointed out, the semantics also often circumscribes the possible intonations, but again only to a limited extent. When speakers of Irish, Welsh, American, or Scottish English read a printed page such as this one there will be differences in their intonation patterns that cannot be ascribed to anything written down. But are these differences part of their language, or do they convey only sociolinguistic information about the speakers?

It is worth considering what kind of phonological theory we would need if we limited ourselves to accounting for the linguistic information that is conveyed by a written language with a good orthography (i.e. a written language such as Finnish or Swahili in which there are few letter-to-sound ambiguities such as written English "read" which can be [rid] or [rɛd]; and no sociolinguistic variations, such as British English "colour" and American English "color"). It is difficult to define the linguistic information conveyed by such a written language in positive terms, but we could say that it is all the encoded aspects of speech except those that convey information about the speaker's identity, attitude, emotions or sociolinguistic background, in so far as these are not conveyed by syntactic word order or lexical devices. This last proviso is especially necessary if we are to include (as most linguists would) some but not all patterns of intonation within phonology.

The role of intonation is undoubtedly the most problematic part of this proposal for phonology. The formulation suggested above is designed to include differences in intonation such as those between statements of the form "That is a cat" and questions such as "Is that a cat?" But it would relegate to a difference in attitude some things that can be expressed in writing such as the incredulous question "That is a cat?" In other words it would consider as <u>linguistic</u> only those intonation patterns that had syntactic or lexical correlates.

Past work on intonation is the only extensive body of phonological work that is not in fact confined to a spoken equivalent of the written language. Thirty years ago, linguists used to argue whether four pitch levels and a number of junctures were sufficient to capture all the meaningful contrasts in English (see, for example, Trager & Smith 1951, Stockwell 1960). Similar discussions are still in progress using a different set of phonetic devices. But in all these discussions of intonation the notion of a meaningful contrast is not the same as it is in discussions of other aspects of phonology. It includes aspects of the emotional functions of speech that are conveying the speaker's attitude. Phonology is not usually considered to include comparable aspects of segments, such as lengthening to indicate superlatives [ɪt wəz bɪ::g] [hi wəz gr:eɪt]. (A

possible exception is Prince (1980), which has a brief discussion of the realization of emphasis in Finnish.) It would seem appropriate to constrain studies of intonation to those aspects of speech that convey simply linguistic information.

If we limit phonology in this way the relation between phonological and phonetic units becomes much more straightforward. The phonologist is no longer under an obligation to describe the phonetic details that characterize the sounds of one language as opposed to another. Consequently there is no need for a complex feature system. As it is so obvious that languages differ in ways that are not used to differentiate phonological oppositions we should clearly give up trying to devise a feature system that can characterize differences between languages. Features should simply be distinctive.

In fact, this view of phonology is fairly similar to the early Jakobsonian view. Thus Jakobson, Fant and Halle (1952) and Jakobson and Halle (1956) were plainly striving to minimize the number of distinctive features needed to account for phonemic oppositions in the languages of the world. They were willing, for example, to subsume under the one feature, Tense/Lax, four consonantal features listed by Trubetzkoy - "the tension feature, the intensity or pressure feature, the aspiration feature and the preaspiration feature" (Jakobson and Halle, 1956:28), on the grounds that no language uses these phonetic possibilities independently. Their emphasis was on what was distinctive within a language. This is very different from the theory propounded by Chomsky and Halle (1968), who do not use the term distinctive features, but instead emphasize that features reflect general phonetic capabilities.

The only problem with a phonology of the kind that concerns itself simply with the patterns of linguistic distinctions is that it is largely untestable. If I say that a certain opposition in a given language is describable in terms of the feature fortis/lenis, and you prefer the feature voiced/voiceless, there is no way of deciding which of us is right. In fact, there is no way of deciding whether any one set of distinctive features is preferable to any other. All one can do is appeal to traditional scientific criteria, such as the parsimony, the observational adequacy and the explanatory elegance of a description. As Murdoch (1981:6) somewhat obscurely comments "Linguistic idealism (is) a dance of bloodless categories."

The view that phonology should not consider sociolinguistic information is, as I have noted, not that theoretically held by most phonologists. For example, Schane (1973), in his statement of how phonetics is part of phonology, says that: "Linguistically significant differences are those which characterize native control of a language." It is also different from the view that I have myself expressed elsewhere: "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" (Ladefoged 1980). But it is an option, and one that represents the practice (but not the stated purpose) of probably the majority of phonologists. Its particular advantage is that it does not require phonologists to think very deeply about phonetics, or about the realization of phonological features. It would make the difference between one dialect and another, or one language and another, a part of sociology, describable (probably in traditional phonetic terms) in the same way as any other indexical behavior, such as the dress, appearance, or patterns of belief that characterize a particular group. Such things are part of culture; and as even Stalin (1950) knew: "Linguistics is not to be confused with culture." It would also regard

phonetic differences conveying emotion, or those aspects of the speaker's attitude that cannot be expressed by syntactic or lexical devices, as part of the subject matter of psychology. In this view, phonologists are left with their own special field: describing and explaining in terms of general linguistic principles the patterns of speech that can convey what we have defined as objective linguistic information. If phonologists are not prepared to consider the phonetic realization of phonological units in detail, this is all phonology could be.

Finally, I might mention another speech gesture that is not used distinctively in any language, the possibility of speaking with tongue in cheek; and I will admit to something of the written equivalent (pen in word processor?) in this paper. I think that the arguments I have proposed for an alternative view of phonology are plausible. But accepting them goes against all my previous endeavours, and I am not yet prepared to regard phonetic detail as not part of phonology.

References

Anderson, S. R. 1978. "Tone features." Tone: a Linguistic Survey. (ed. V. Fromkin), 133-176. New York: Academic Press.

Bolinger, D. 1977. "Another glance at main clause phenomena." Language 53.3, 511-520.

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

Halle, M. and Stevens, K. 1971. "A note on laryngeal features." MIT RLE Quarterly Progress Report 101, 198-213.

IPA. 1979. The Principles of the International Phonetic Association. London: University College.

Jakobson, R., Fant, C.G.M. and Halle, M. 1951. Preliminaries to Speech Analysis. MIT Acoustics Laboratories Technical Report, 13.

Jakobson, R. and Halle, M. 1956. Fundamentals of Language. The Hague: Mouton.

Ladefoged, P. 1967. Three areas of experimental phonetics. London: Oxford University Press.

Ladefoged, P. 1980. "What are linguistic sounds made of?" Language 56, 485-502.

Ladefoged, P. and Bhaskararao, P. 1983. "Non-quantal aspects of consonant production: a study of retroflex consonants." J. Phonetics 11, 291-302.

Ladefoged, P. and Wu, Z-J. 1984. "Places of articulation: an investigation of Pekingese fricatives and affricatives." J. Phonetics 12.

Lerner, A.J. and Loewe, J. 1956. My Fair Lady. New York: Coward-McCann.

Lindau, Mona. 1982. "Phonetic differences in glottalic consonants." UCLA Working Papers in Phonetics. 54, 66-77.

Maddieson, I. 1984. Patterns of Sounds. Cambridge: Cambridge Univeristy Press.

Murdoch, I. 1980. Nuns and Soldiers. Harmondsworth: Penguin Books.

Nihalani, P. 1983. "Voice quality and its implications." Abstracts of the Tenth International Congress of Phonetic Sciences 743. Dordrecht, Holland: Foris.

Porter, C. 1936. Anything Goes: A Musical Comedy. New York: Harms.

Prince, A. 1980. "A metrical theory for Estonian quantity." Linguistic Inquiry 3, 511-562.

Schane, S. 1973. Generative Phonology. Englewood Cliffs, N.J: Prentice-Hall.

Shaw, G.B. 1920. Pygmalion: a Romance in Five Acts. London: Constable.

Stalin, Josef V. 1950. "Marxism and problems of linguistics." Pravda June 20 1950. Moscow.

Stockwell, Robert P. 1960. "The place of intonation in a generative grammar of English." Language 36, 360-367.

Sweet, H. 1890. A Primer of Phonetics. Oxford: Clarendon Press.

Trager, G. and Smith, H.L. 1951. An Outline of English Structure. Norman, Oklahoma.