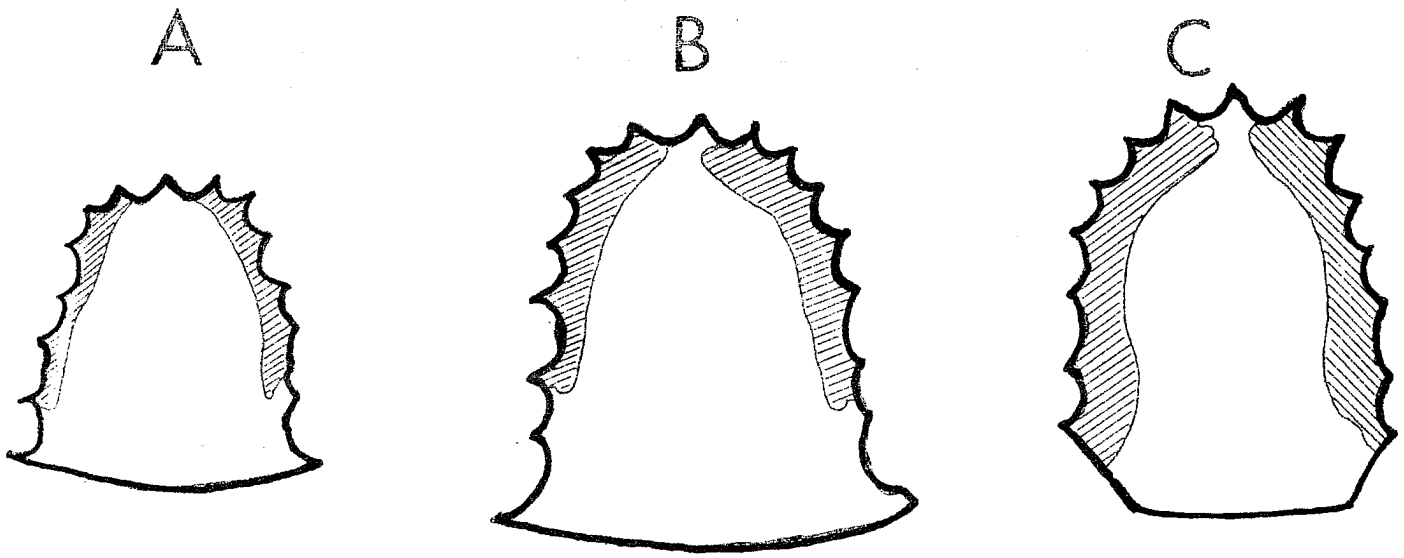
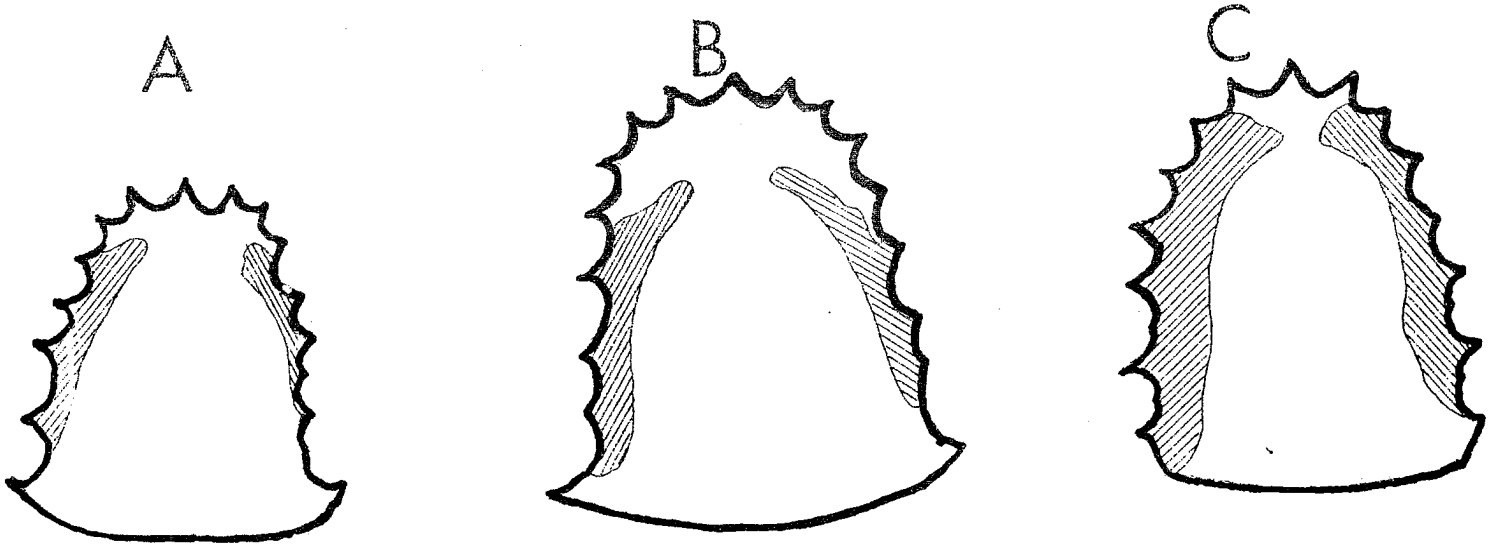
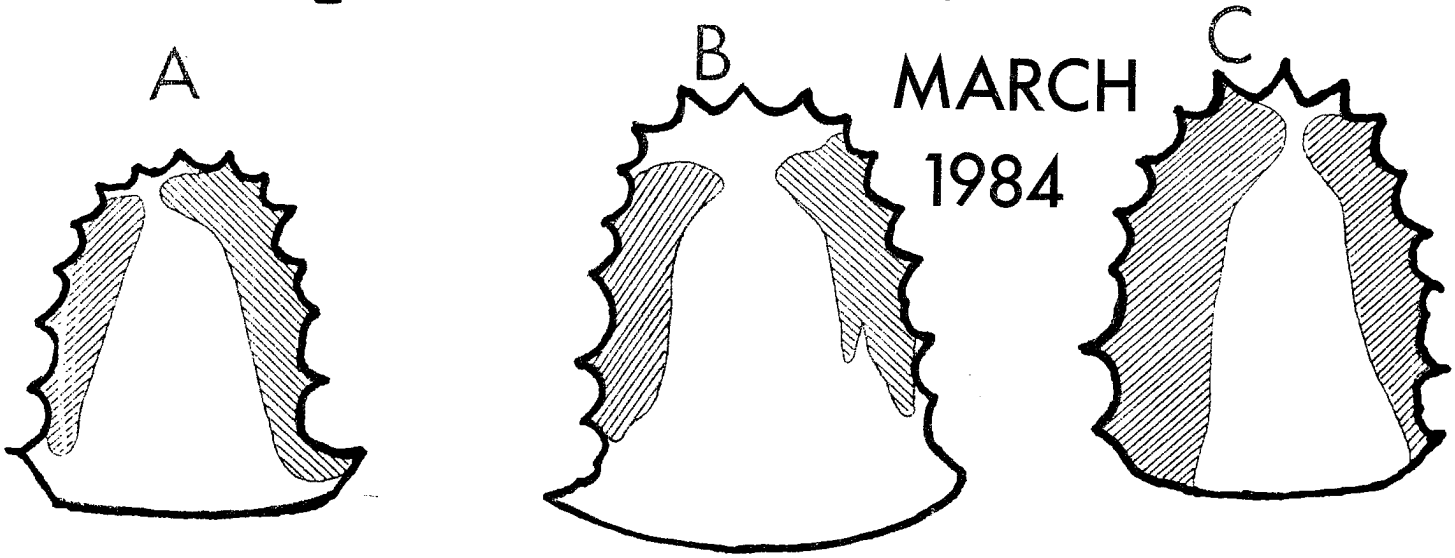


ucla working papers



in phonetics 59



MARCH
1984

UCLA Working Papers in Phonetics 59

March 1984

Testing an aerodynamic model with measured data from Korean

Sarah N. Dart

I. Introduction

In the languages of the world various states of the glottis are exploited to distinguish sounds which otherwise differ very little in articulation. Voiced segments are contrasted not only with voiceless segments, but also with segments utilizing different types of voice such as creaky voice in Hausa and breathy voice or murmur in some of the Indo-Aryan languages in the Indian subcontinent. In addition to this, in some languages stops are said to be distinguished by degree of muscular tension in the vocal folds or by the timing of the glottal opening and closing gesture. In order to find out in more detail how all these distinctions are made, phoneticians have developed new techniques of investigation. One such technique is a method for measuring oral air flow and intraoral air pressure during the production of contrasting segments, which will be described in detail below (see also Javkin and van der Veen 1983). Since both oral pressure and flow are influenced by a change in glottal adjustment, one can infer a certain amount about the glottal state by measurement of these parameters. These inferences must be made, of course, with some caution, since other factors besides glottal state can influence oral pressure and flow (for example, changes in subglottal pressure, tension in the vocal tract walls, active expansion of the vocal tract, etc.). If used in conjunction with other methods of investigation, air pressure and flow data can help give a more complete picture of sound production. In the present study, pressure and flow measurements were taken of the Korean "fortis" and "lenis" stops. These results will be discussed with reference to supplementary information gained from the use of a computer implemented aerodynamic model which will be described in detail in a later section.

II. Experimental Apparatus

Oral (and nasal) airflow was recorded using a modified respiratory mask with a fine stainless steel gauze which exhibits a known amount of resistance through which the outgoing air must pass (as described by Rothenberg 1973). The flow was calculated from the pressure difference across the gauze. A Gaeltec catheter tip pressure transducer was inserted through a hole in the mask designed for that purpose. The tube containing the pressure transducer was placed between the lips of the subject in such a way that when the lips were closed, the transducer measured the air pressure inside the mouth, without touching the walls of the oral cavity and without being hit by any of the moving articulators (for this reason the consonants were limited to bilabials). After each session the air pressure and flow devices were calibrated, the pressure by the use of a standard U-tube manometer and the flow by introducing a known voltage (since the flow mask is regularly checked to insure that it registers .98 volts for a flow of 1000 ml/sec.).

The pressure and flow were monitored on an oscilloscope and recorded on an oscillomink inkwriter. As an aid to segmentation, an audio recording was made simultaneously on another channel. In order to get a higher quality acoustic record (since the mask muffled the sound somewhat) a separate recording was also made in a sound-proof booth of the same words. Figure 1 is a schematic representation of the experimental procedure.

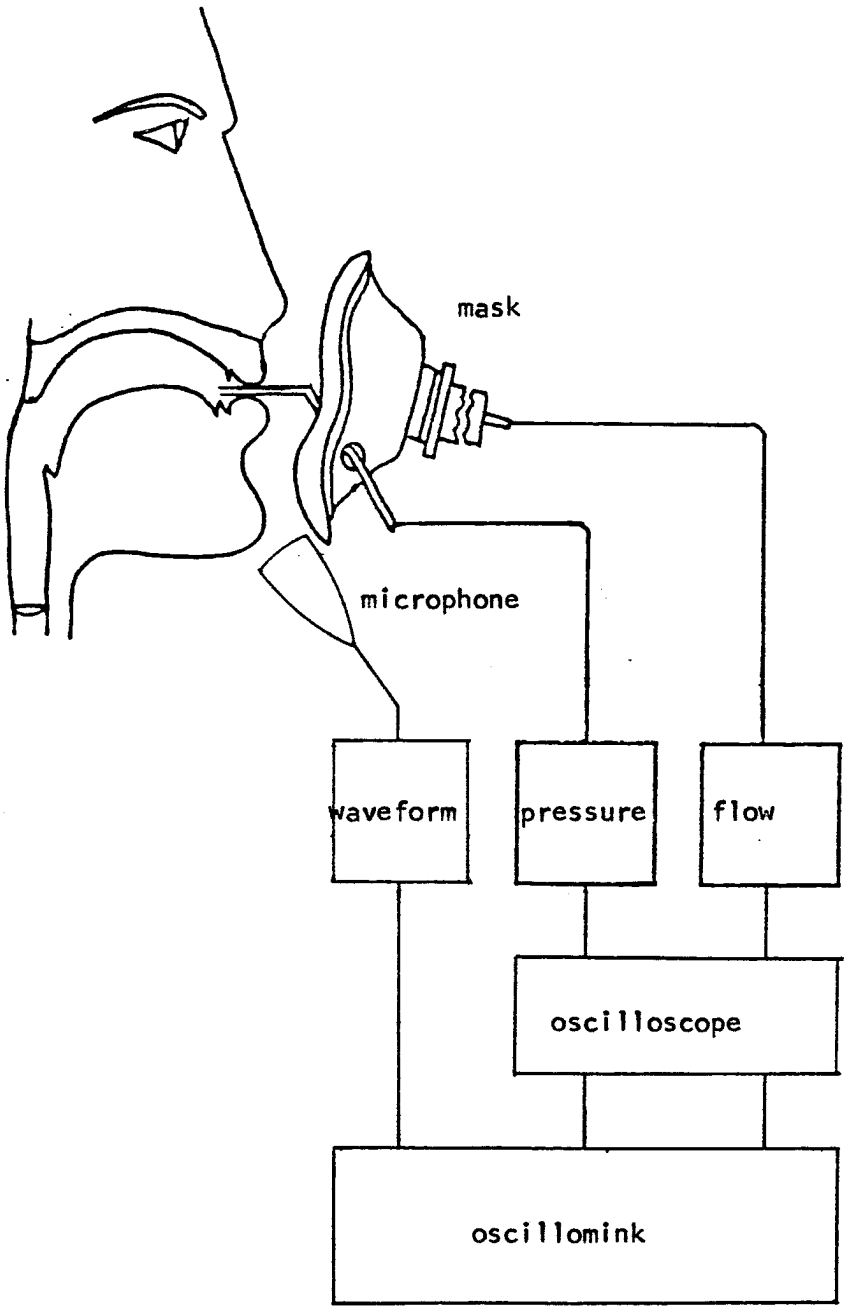


Fig.1. Diagram of experimental apparatus.

In addition to these recordings, an F-J Electroglottograph was used on one of the Korean subjects, from which the pitch was also calculated. The electroglottograph consists of two small metal plates attached to a strap around the subject's neck. The plates were positioned on either side of the larynx and a small electric current was passed between them. From variation in the resistance, it can be inferred to what degree the vocal folds are adducted (since when they are together, the electric current passes through easily).

III. Korean

Korean has a three-way distinction between voiceless stops in initial position. These are exemplified in the following minimal triplet: [p*aŋ] "bread", where the initial stop is said to be unaspirated and "tense" or "fortis"; [paŋ] "room" with a "partially aspirated" (or "lenis") stop; and [phaŋ] "bang ", where the stop is heavily aspirated. At first glance this would seem to be a simple difference in voice onset time and certainly the heavily aspirated stop may be distinguished on this basis alone. The other two stop types, however, have been shown in some studies (Kim 1967, Abramson and Lisker 1971, Han and Weitzman 1970) to have overlapping VOT values. But even on occasions when the VOT values are identical, the two stops are still distinguishable to native speakers, as suggested by the results of a perception test conducted by Han and Weitzman (1970). Clearly, VOT differences are not enough to distinguish what will henceforth in this paper be called "fortis" and "lenis" stops.

Previous studies

One of the first investigators to examine Korean stops experimentally was Chin-Wu Kim (1965, 1967, 1970). His cineradiographic study showed greater pharynx width during the occlusion of fortis stops than during lenis stops, although this difference was only convincing for bilabials. Also noted was a slight raising of the larynx during fortis stop closures. Kim hypothesized that both of these effects resulted from heightened subglottal pressure or "stress" in fortis stops, although it is not clear why this would necessarily be the case, since both pharynx expansion and larynx raising may be effected by means of muscular activity completely independent of that involved in raising the subglottal pressure.

Han and Weitzman (1970) also showed that the fundamental frequency at voice onset after fortis stops is generally higher than that after lenis stops. They found also that voice onset after fortis stops was very sharp with relatively undamped harmonic partials in comparison to the lenis stops. This would follow very well from the widely held view that fortis stops are characterized by muscular tension in the vocal folds and/or vocal tract walls.

Hardcastle (1973) referred to "isometric muscular tension", i.e., the tensing of a muscle which does not undergo shortening as a result, present in the vocal folds and the walls of the pharynx during articulation of the fortis stop. He supported his claim by acoustic evidence such as the higher frequency of vocal fold vibration at the onset of voicing following fortis stops (an indication of tension in the vocal folds) and the sharper formant structure and more clearly defined harmonic partials which might indicate general tension in the walls of the supraglottal cavity.

Hirose, Lee and Ushijima (1974) made an important contribution to knowledge of vocal fold activity during Korean stops with their EMG study of some of the

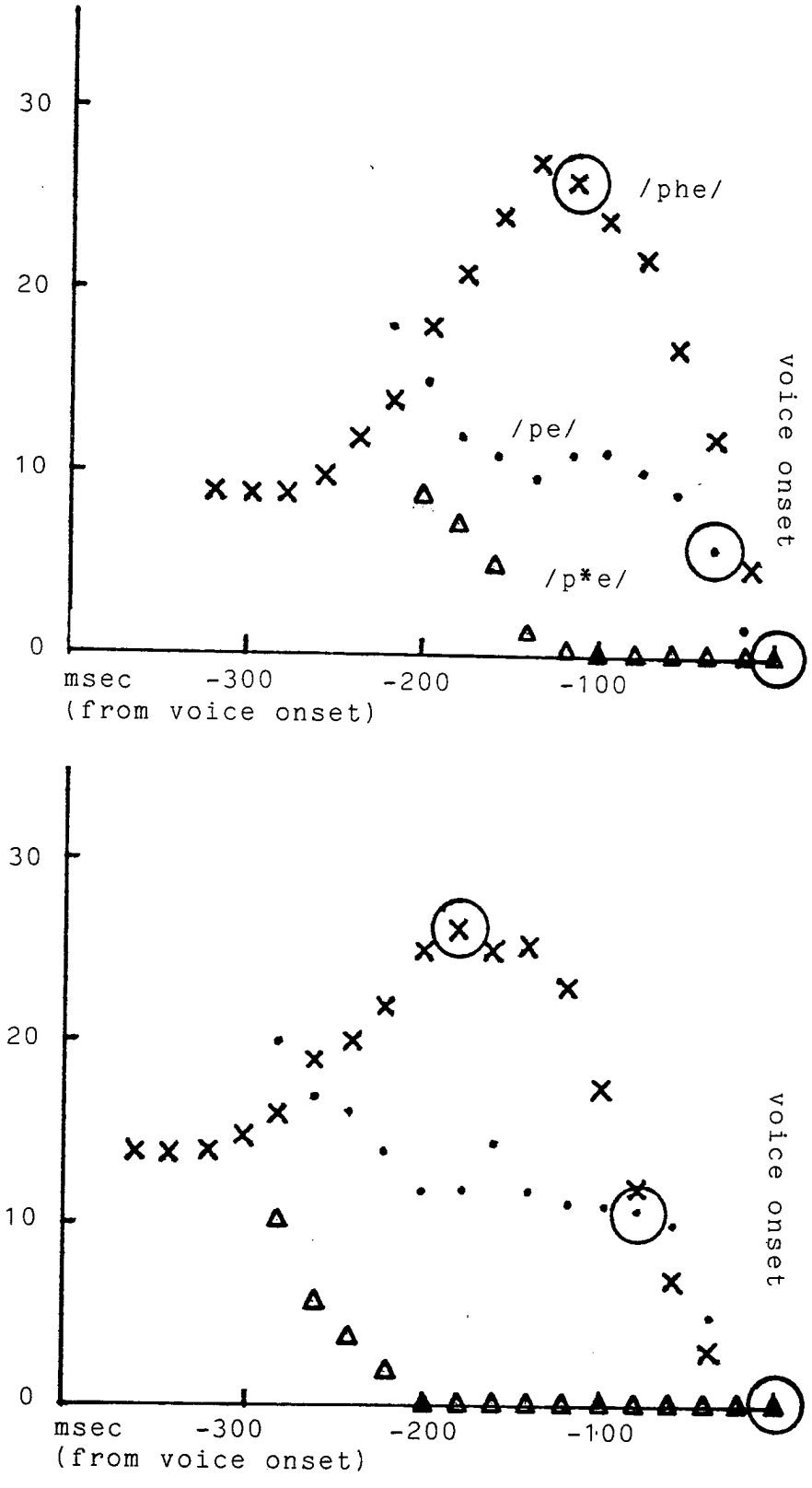


Fig.2. Glottal adjustment on an arbitrary scale for fortis (Δ), lenis (•) and aspirated (X) stops in Korean from Kagaya (1974). Circle indicates articulatory release.

intrinsic laryngeal muscles. Perhaps the most interesting finding was the marked increase in lateral cricoarytenoid and vocalis muscle activity just prior to release in fortis stops, presumably resulting in tension of the vocal folds and constriction of the glottis.

Kagaya (1974) photographed the vocal folds during Korean stop production with the use of a fiberscope, revealing very different laryngeal adjustments for the three stop types. It can be seen in Figure 2 (after Kagaya 1974) that the maximum glottal opening during occlusion was largest for the aspirated stop, intermediate for the lenis stop and least of all for the fortis. The timing of the closing gesture relative to articulatory release also varies between stops. For the aspirated stops, release generally occurred near the moment of maximal glottal opening. With lenis stops, although the glottis was still quite open at release, there was a more or less continuous decline in glottal area throughout the occlusion. During the fortis occlusion, on the other hand, the vocal folds came together well before release.

The major problem with all of the above mentioned studies was the lack of a sufficient number of subjects. One cannot confidently generalize from EMG data on one speaker, fiberscopic data on two others and x-ray studies of yet another single speaker. It is not surprising that some conflicting findings have been reported, since from such a small number of speakers, it is difficult to separate salient generalizations from speaker-specific idiosyncracies. I will discuss some characteristics which appear to vary among speakers in a later section.

Experiment

In the experiment ten native speakers of Korean were recorded, six males and four females, all between the ages of 18 and 26. Seven were speakers of the Seoul dialect, two of the Kangwon Do dialect and one of the Kyonsang Nam Do dialect. Their residency in the United States ranged from 1 to 3.5 years, with the mean being 5.5 years.

Four minimal word pairs were repeated by each subject six times: three times in the order lenis-fortis and three times in reverse order. In the following words the asterisk represents the fortis nature of the stop.

Table 1: Korean Word Pairs

1a) [pjə] "rice"	1b) [p*jə] "bone"
2a) [paŋ] "room"	2b) [p*aŋ] "bread"
3a) [ˊpiə] "empty"	3b) [ˊp*iə] "sprained"
4a) [ˊpɛə] "soak through"	4b) [ˊp*ɛə] "pull out"

A typical example of an oscillomink printout is given in Figure 3. The arrows indicate the points measured. Measurements were made of peak oral pressure during the occlusion and peak oral flow immediately after articulatory release.

Results

Figure 4 represents the air pressure and flow data collected for each word pair. On each graph peak flow on the ordinate (in litre/sec.) is plotted against peak pressure (in cm. H20) along the abscissa. Each point represents an average of the 6 tokens of the word indicated recorded from a single speaker. The solid

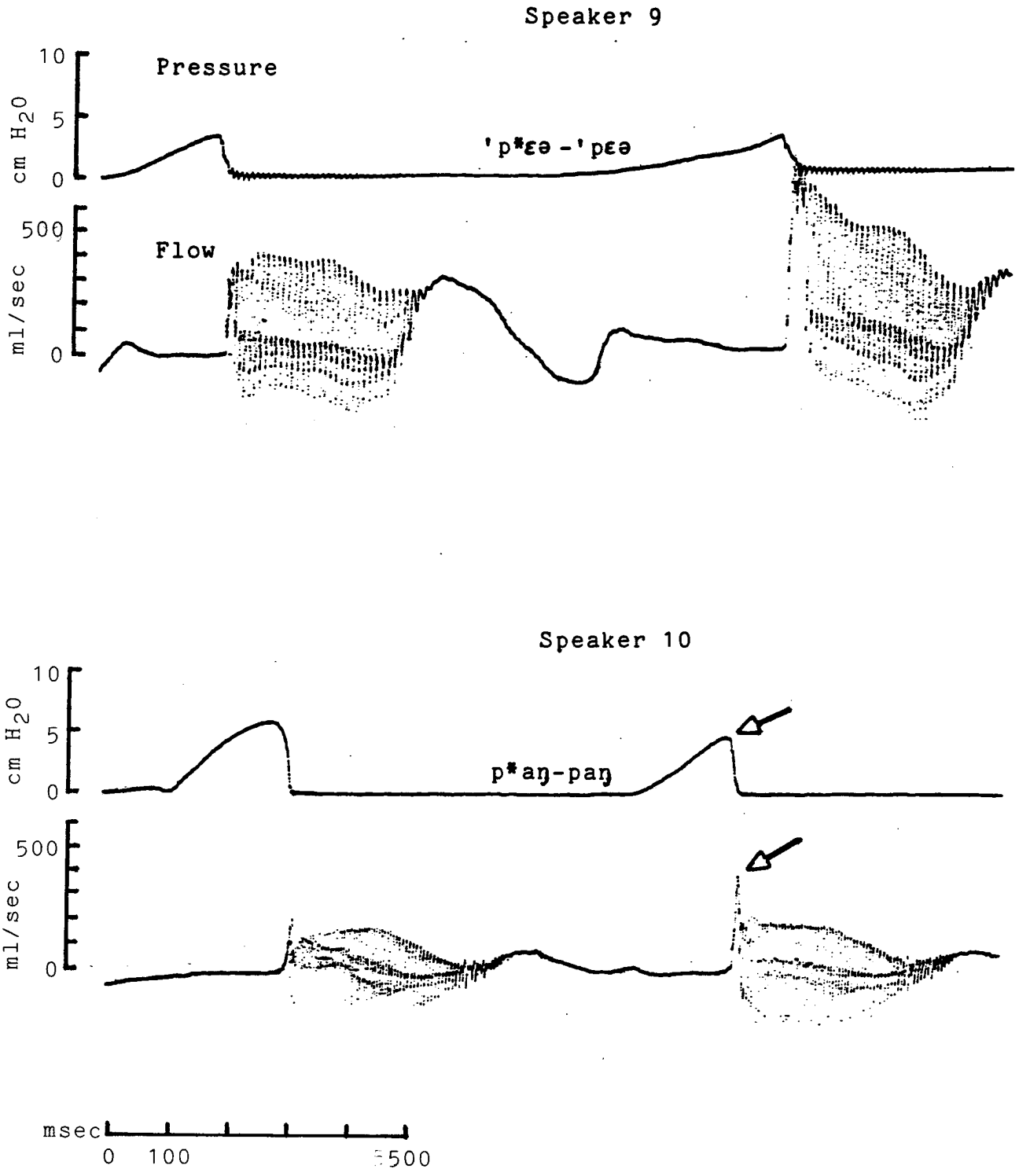


Fig.3. Oscillomink traces of oral air pressure and flow during the production of two different Korean word pairs. Arrows indicate points measured.

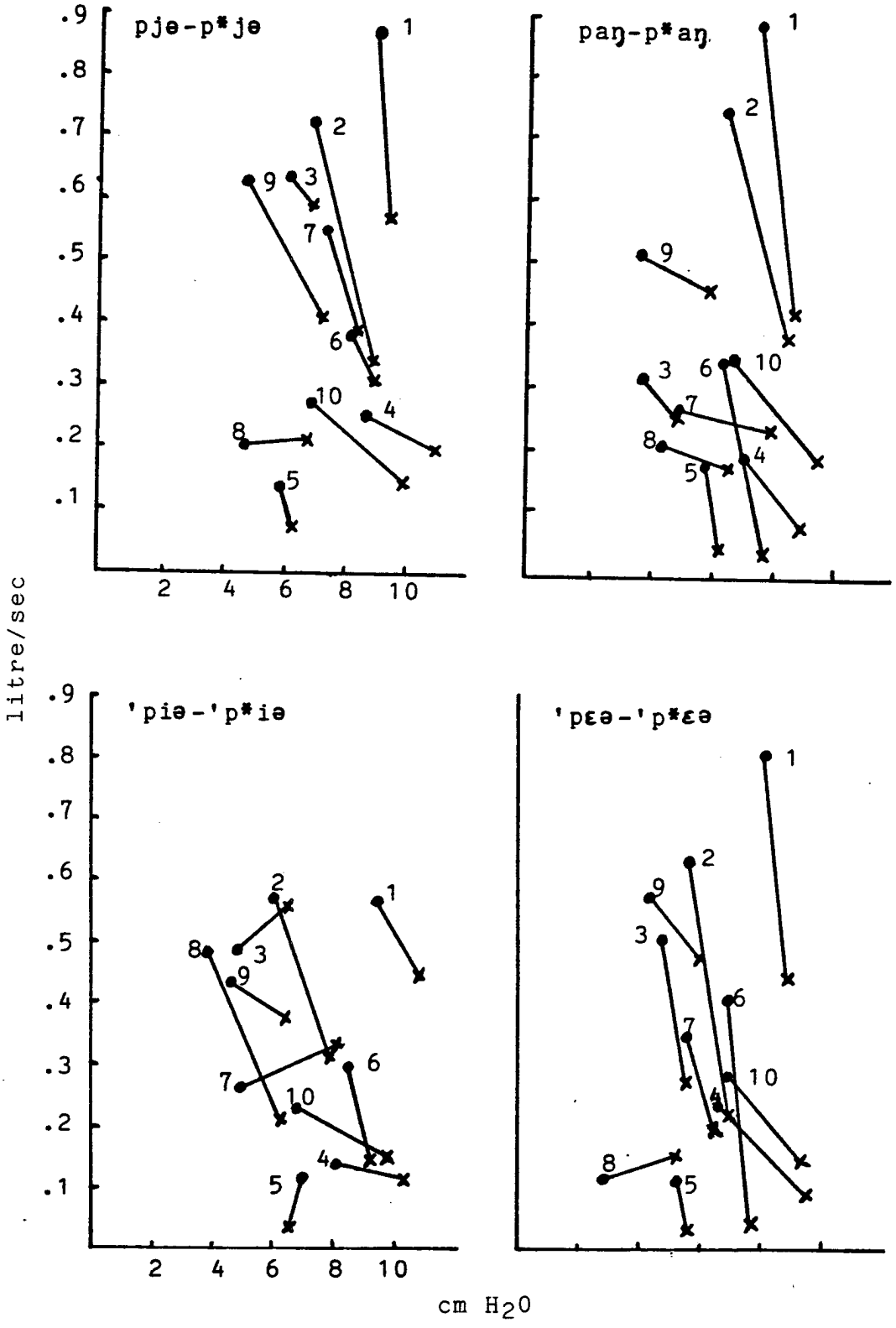


Fig.4. Pressure and flow of lenis (solid points) and fortis(crosses) stops of 10 speakers for each word pair.

points represent lenis stops and the crosses their fortis counterparts. A solid line connects the two points representing the word pair for each speaker.

Discussion

As can be seen in the graphs, there is a great deal of variation among speakers in absolute value of pressure and flow. For each speaker, however, there exists a clear distinction between the two sounds. The overall tendency is for fortis stops to have higher pressure and lower flow values than lenis stops in the same environment. Speakers differ, however, in which parameter has more importance in the distinction. Speakers 1 and 2, for example, appear to make the distinction more on the basis of flow than pressure in most of the word pairs. Speakers such as 4 and 10, on the other hand, have less of a flow difference and more of a difference in oral pressure. These variables also changed according to which word pair was being uttered. Some speakers appear to use different production strategies in different word environments. In a few cases (speakers 3 and 7 in word pair 3 and speaker 8 in word pairs 1 and 4) the usual flow differences are even reversed, giving a higher flow rate to the fortis stop, but the distinction in pressure is maintained as expected.

In all, the word pair which showed the least differentiation between fortis and lenis was word pair 3, ['piə]-['p*iə]. For most of the speakers the reason seems to lie in an unusually low flow for the lenis member of the pair. We could hypothesize that this is due to the following stressed high front vowel. It has been claimed that high vowels have more oral volume than low vowels because of greater pharyngeal width (Smith and Westbury, 1975). This would mean greater surface area in the vocal tract and a corresponding decrease in the initial flow rate at release, since more flow would be absorbed by the elasticity of the vocal tract wall tissue. Another factor which might influence flow rate is the degree of lip opening which is smaller in high vowels than in low. A smaller oral constriction would allow less air through and consequently decrease the flow rate after release. The same effect could be attributed to the narrow palatal tongue constriction in the high front vowel. A combination of these factors could decrease the flow in this particular word pair. We might also expect to see a lowered flow in the lenis member of word pair 1 [pjə] - [p*jə] because of the presence of the high front glide [j]. In fact, looking at the average difference in flow values for the ten speakers in each word pair, it is clear that this word pair is similarly affected, but to a lesser degree.

Table 2: Average difference in flow values between fortis and lenis members of the four word pairs.

Word Pair	1	2	3	4
Flow diff. (lenis minus fortis in ml/sec)	139	171	88	189

Word pairs 1 and 3 have less of a flow difference than 2 and 4. The difference is least in word pair 3 presumably because the high front vowel is stressed, whereas in word pair 1, the glide is of too short a duration to have as much effect. The problem of why the flow decreases should show up mainly after the lenis stops will be discussed in a later section.

Aerodynamic Model

In order to understand the possible articulatory and glottal differences in stop production which result in the observed pressure and flow differences, an aerodynamic circuit model was used. Using such a model forces one to consider every variable in determining input values and helps to narrow down the possibilities for realistic interpretations of the data. Once set up for the known values, the model can serve as a testing ground for hypotheses concerning the less well understood components of an articulation.

The aerodynamic model used is a computer implemented electrical analog derived from Rothenberg (1968), similar to the model described by Müller and Brown (1980) and Westbury (1983). Voltage is taken to be the analog of air pressure and current is the analog of volume velocity airflow. The model gives as its output oral pressure, subglottal pressure, flow through the glottis and flow through the mouth opening. For simplicity in calculation, some of the input parameters are regarded as invariant during a given simulation, including some of the glottal dimensions, oral tract wall impedance, vocal tract volume and surface area. Other input parameters may vary over time, notably respiratory muscle force, distance between the articulators, distance between the vocal folds and active expansion of the supraglottal cavity.

The model was used to try to simulate the observed pressure and flow data reported above. Input values for the present study were estimated from previously mentioned results reported in the literature of fiberoptic and x-ray studies of glottal opening, x-ray studies of larynx height and pharynx width and from acoustic measurements of VOT values for the Korean fortis-lenis distinction. In order to get a realistic input for the model, the closure durations were measured from the data of all ten speakers. Closure was considered to begin as the oral pressure curve began to rise and to end at the beginning of flow rise at release. It was found that the fortis closures were considerably longer than the lenis closures. This result did not differ significantly between word pairs. The average closure duration of all lenis stops was 133.5 msec and that of all fortis stops was 188.25 msec. For modeling purposes the lenis stop was given a duration of 135 msec and the fortis 190 msec. Each stop was given a 20 msec oral release gesture from fully closed to fully open values.

Figure 5 illustrates the input and output values for the model, given the above difference in closure duration and the difference in glottal area function as estimated from Kagaya's (1974) data. It is generally agreed that during fortis stops the vocal folds come fairly tightly together well before articulatory release. After release the folds gradually return to a position closer to that of normal voicing. The increase in vocal fold tension was modeled, as can be seen on the left side of Figure 5, by a narrowing of the distance between the vocal folds. The model has no parameters directly reflecting vocal fold tension. It was presumed that greater tension would have the effect of bringing the vocal folds closer together and that by this substitution, similar results (i.e., decreased flow through the glottis) would be obtained. The hypothesized tension in the vocal folds is also supported by the results of the glottograph pitch record in the present study which showed consistently a higher pitch immediately after release of the fortis stop. On the righthand side of the figure, the glottal area function for the lenis stop is given. Not only is the initial glottal opening greater than that during the fortis stop, but the vocal folds do not become adducted for voicing until approximately 40 msec after release.

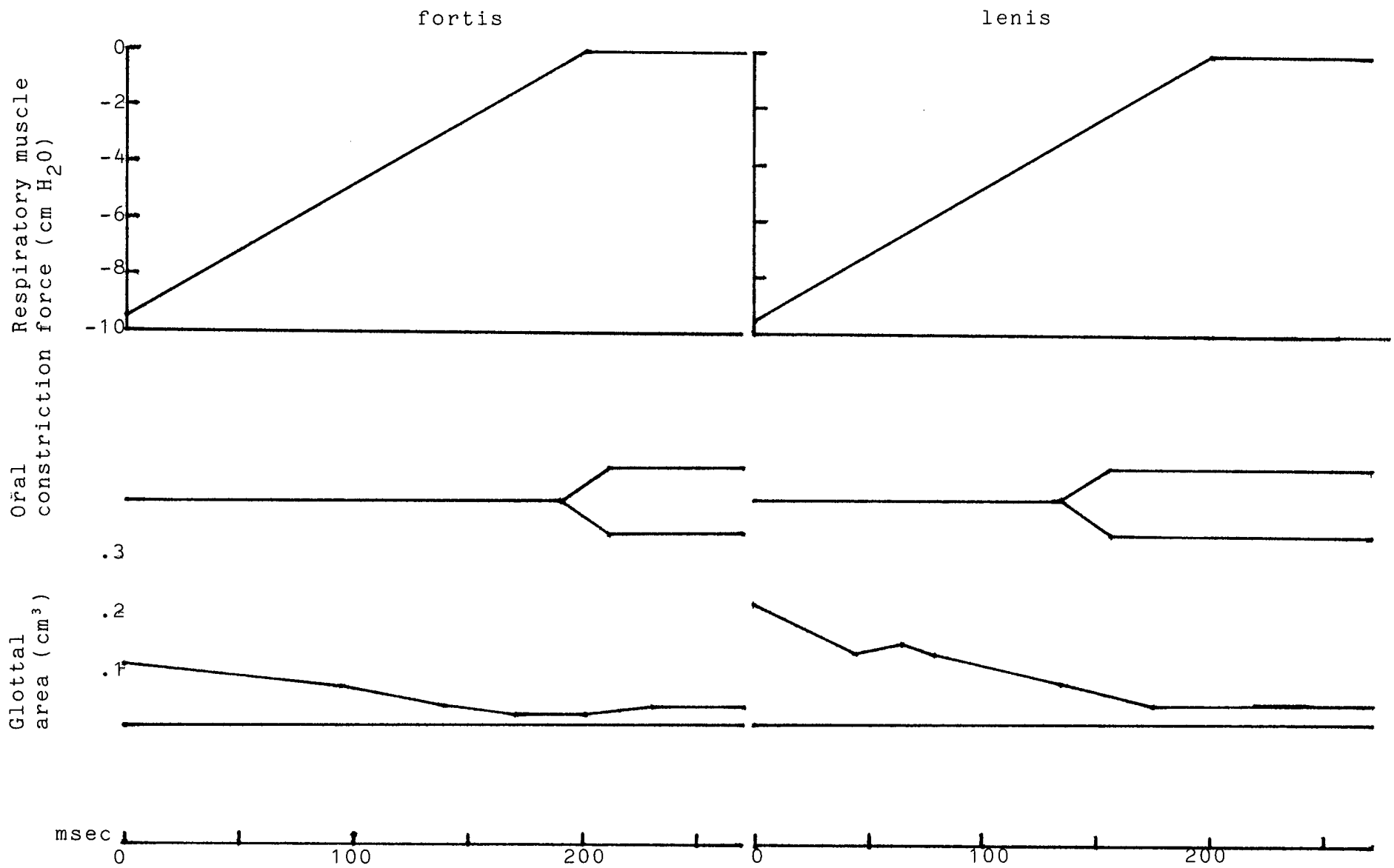


Fig.5. Input values for the aerodynamic model simulating measured glottal area and closure duration differences for fortis and lenis stops.

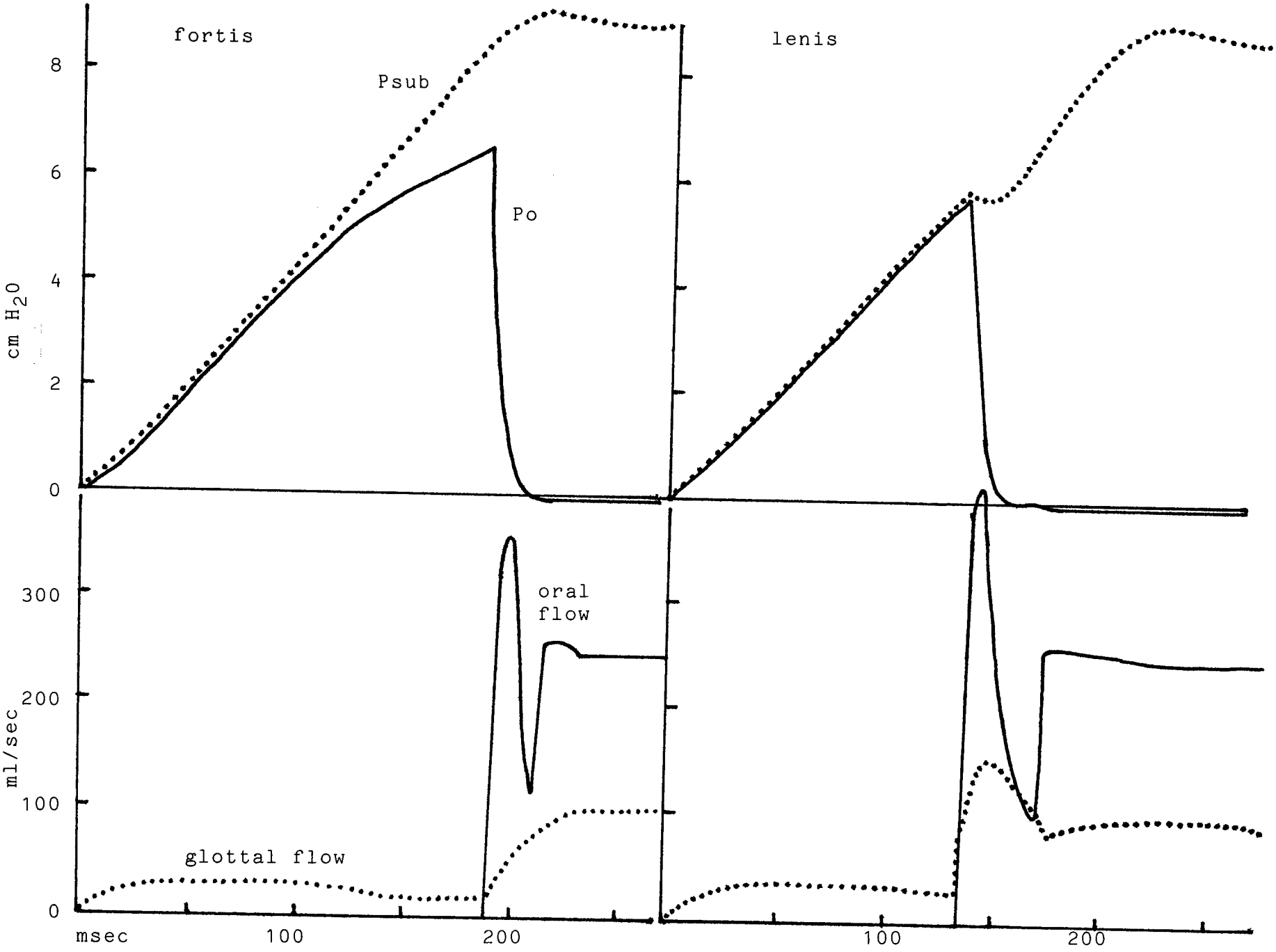


Fig.6. Output values from aerodynamic model inputs given in Fig.5.

If these differences in glottal area function and closure duration are modeled without changing any other variables, that is, with identical values of vocal tract wall tension, the pressure and flow produced are as shown in Figure 6, with peak values as follows:

lenis P	5.7cm H20	fortis P	6.6cm H20
F	411ml/sec	F	357ml/sec

The average pressure difference between lenis and fortis stops for all word pairs and all speakers was 1.4cm H20. The average flow difference was 140ml/sec. The results of the simulations in Figures 5 and 6 show a pressure difference of .9 cm H20 and a flow difference of 54ml/sec, unrealistically low figures for both values.

Kim (1965), Hardcastle (1973), and others have suggested that the fortis stop is characterized by greater muscular tension in the vocal tract walls. Figure 7 shows the results of modeling such a tension difference. The output values of pressure and flow for the lenis stop are now 5cm H20 and 486ml/sec respectively and for the fortis stop are 6.9cm H20 and 300ml/sec, showing that a greater degree of vocal tract wall tension leads to an increase in oral pressure and a decrease in peak flow value. The pressure increase can be understood as a result of stiffening the walls, which is effectively the same as decreasing supraglottal cavity volume, since it calls for a reduction in the possibility of passive vocal tract expansion. This decrease in elasticity of the cavity walls also contributes to a lower peak flow by decreasing the amount of elastic recoil of the walls and thereby slowing down the initial flow velocity at release. The resulting pressure and flow differences shown in Figure 7 are 1.9cm H20 and 186ml/sec, corresponding fairly closely to the averaged values from the data.

In addition to the differences in peak values, however, a distinct difference in oral pressure curve shape was also noted between the two stop types, as can be seen by referring back to the oscillomink traces given in Figure 3. Not only was pressure during the fortis stop generally higher than that of the lenis, but the top of the fortis curve was rounded before release, while pressure during the lenis stop went up linearly to a point at release and then dropped off abruptly. The combination of the timing differences in the release of the oral constriction, plus glottal area and vocal tract wall tension differences did not produce these differences in oral pressure curve shape, and accordingly, some other input changes must be postulated.

Kim (1965), Kagaya (1974) and others have hypothesized heightened subglottal pressure during the fortis stops. This could be achieved either by lowering the larynx, or by a more rapid increase in respiratory muscle activity. Figure 8 shows the results of the latter strategy, with a more rapid increase in respiratory muscle force during the first 150 msec of the stop. The resulting curve resembles that of Speaker 9 in Figure 3.

Larynx lowering, on the other hand, shown on the left side of Figure 9, caused a lowering of the pressure curve just before release which was more typical in speakers such as Subject 10 in Figure 3. These inter-subject differences agree with results given in the literature, where some subjects exhibited larynx lowering before the release of fortis stops, others a slight raising of the larynx and yet others, no apparent larynx movement at all (see Kim (1967), Umeda and Umeda (1965) and Kagaya (1974)).

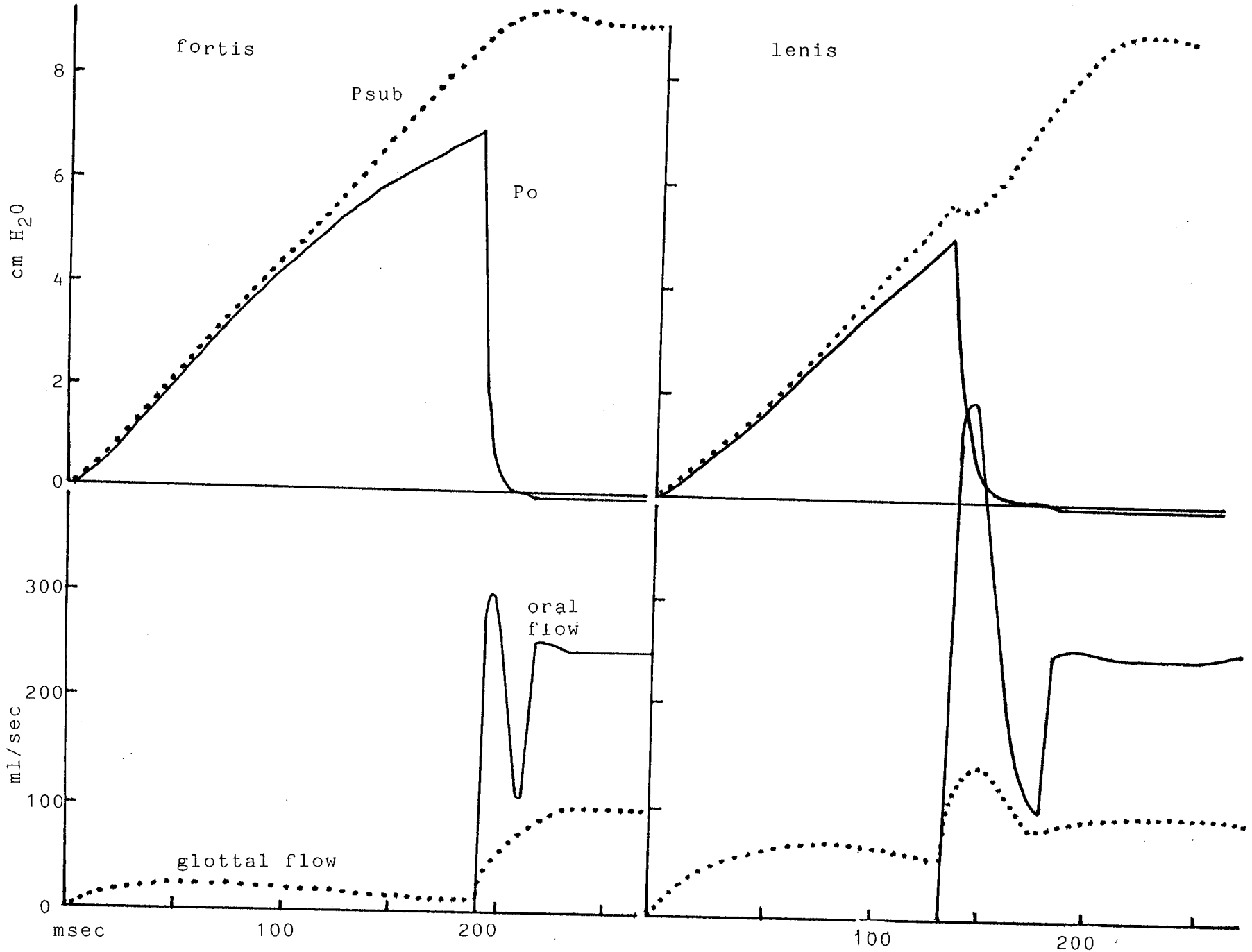


Fig.7. Output values from aerodynamic model inputs given in Fig.5. with the addition of lax vocal tract wall values to the lenis stop and tense wall values to the fortis stop.

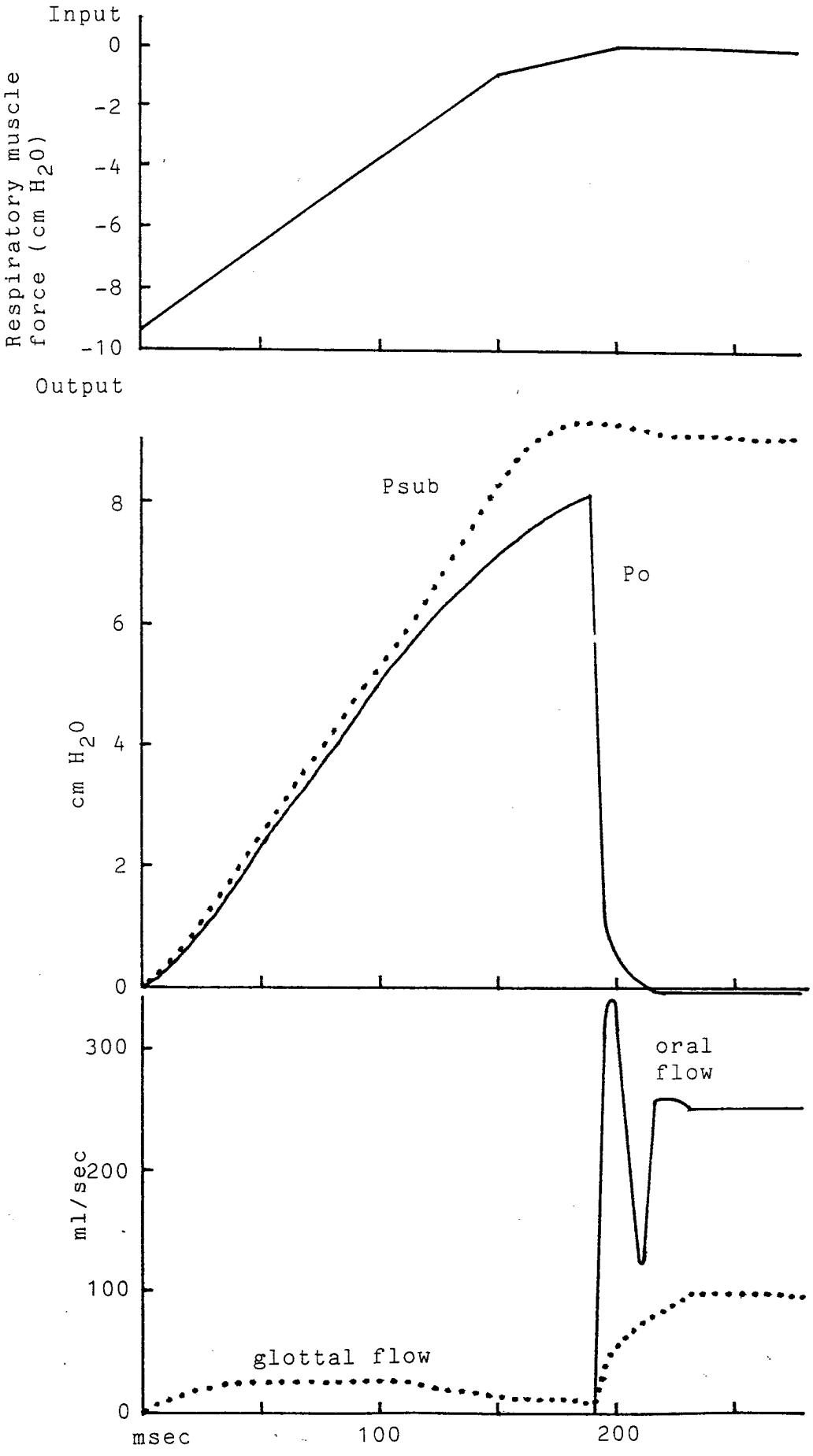


Fig.8. Simulated fortis stop as in Fig.7, but with a more rapid increase in respiratory muscle force as shown.

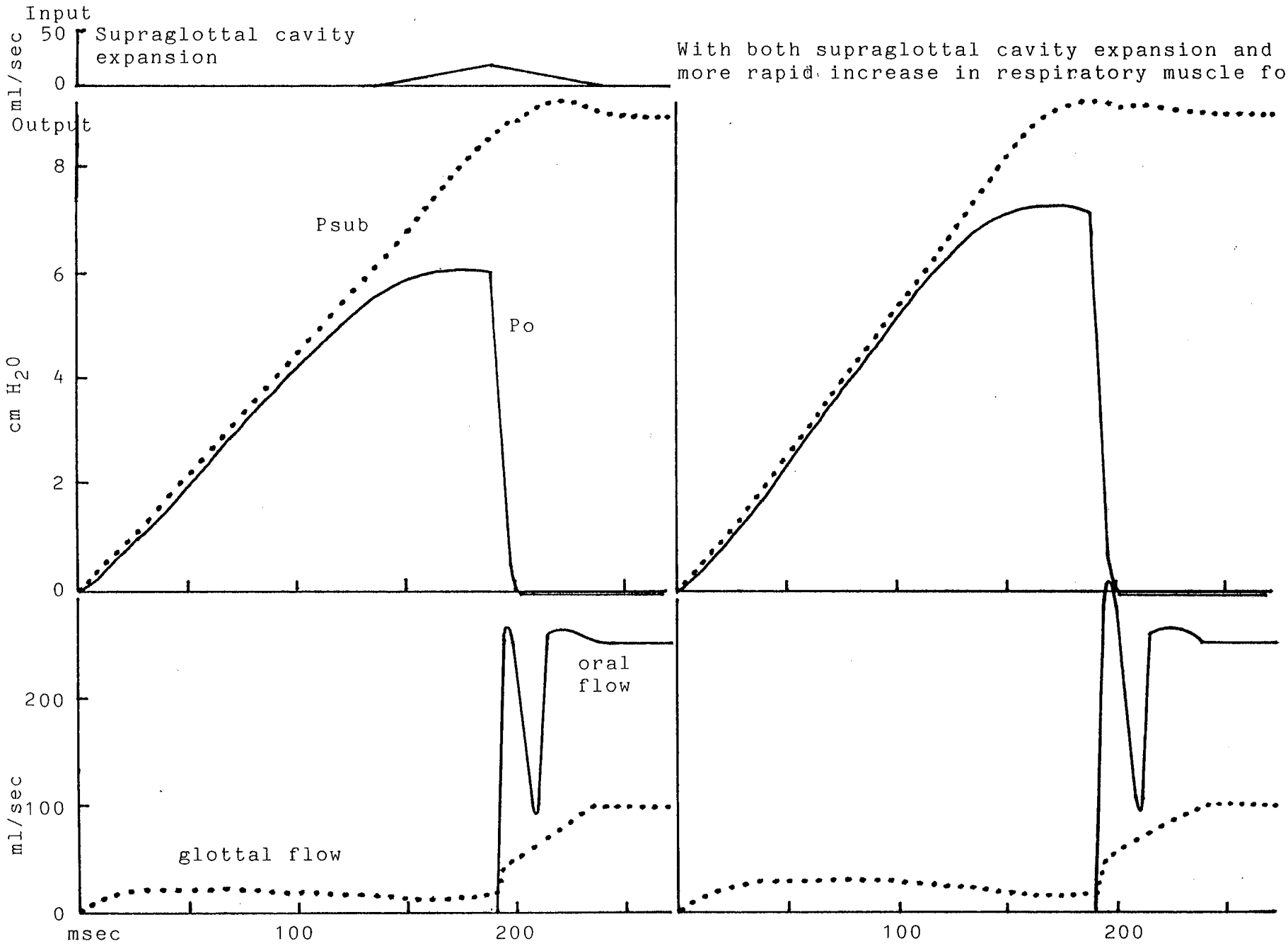


Fig.9. Outputs of simulated fortis stops with supraglottal cavity expansion (left) and both cavity expansion and a more rapid increase in respiratory muscle force (right).

Independent of curve shape, the oral pressure value is proportionately too high in the simulation showing a more rapid increase in respiratory muscle force (Figure 8) and too low when an increase in supraglottal cavity volume is modeled (left side of Figure 9). On the right side of Figure 9, both of these strategies are combined, giving acceptable average pressure and flow values.

It was noted in an earlier section that speakers seemed to differ in whether they made the lenis-fortis distinction more on the basis of pressure differences or flow differences. The output of the simulation just discussed gives peak values of 7.2 cm H20 oral pressure and 302 ml/sec flow for the fortis stop which, when compared with the lenis values in Figure 7 (5 cm H20 oral pressure and 484 ml/sec flow) gives a pressure difference of 2.2 cm H20 and a flow difference of 184 ml/sec. These values correspond more to an average over all the speakers than to any one speaker of the group. The previously discussed simulations suggest the factors that could be involved in speaker specific production strategies. The rapid increase in respiratory muscle force (Figure 8) gave a large pressure difference, whereas the expansion of the supraglottal cavity (Figure 9, left) resulted in a large flow difference and a small difference in oral pressure.

Also noted earlier was the lowered flow in the lenis member of word pair 3 [ˈpiə]-[ˈp*iə]. By simulating a high front vowel following the stop (i.e., modeling an increase in supraglottal cavity volume and a decrease in the area of oral constriction representing the lip opening) a decrease in flow was observed in both the lenis and fortis stops. The decrease was slightly greater, however, in the case of the lenis stop. This is perhaps due to the fact that the lenis stop was modeled with lax vocal tract wall values which, with the increase in supraglottal cavity volume, gave effectively a greater volume increase because of the elasticity of the walls.

To summarize, Korean fortis stops are characterized by higher oral pressure and lower oral flow than their lenis counterparts. This is due in part to the closely adducted vocal folds in the fortis stop. In addition, evidence from modeling leads us to postulate tenser vocal tract walls for the fortis stop and a more rapid increase in respiratory muscle force. For some speakers larynx lowering or other supraglottal cavity expansion also appears to occur just before release of the fortis stop.

Acknowledgments

This paper is a portion of my Master's Thesis, a shortened form of which was presented at the 106th meeting of the Acoustical Society of America, November, 1983. The work was supported by USPHS grant NS 13163-02 to Peter Ladefoged.

References

Abramson, A.S. and Lisker, L. (1971). Voice timing in Korean Stops. Status Report on Speech Research, No. 27, Haskins Laboratories New Haven.

Bell-Berti, F. (1975). Control of pharyngeal cavity size for English voiced and voiceless stops. JASA 57 No.2:456-461.

Han, M.S. and Weitzman, R.S. (1970). Acoustic features of Korean /P,T,K/, /p,t,k/ and /ph,th,kh/. Phonetica 22 No.2:112-128.

Hardcastle, W. J. (1973). Some observations on the tense-lax distinction in initial stops in Korean. Journal of Phonetics 1: 263-272.

Hirose, H., Lee, C.Y. and Ushijima, T. (1974). Laryngeal control in Korean stop production. Journal of Phonetics 2:145-152.

Javkin, H. and van der Veen, R. (1983). A portable phonetics laboratory. Paper given at 106th meeting of the Acoustical Society of America Meeting, San Diego, CA.

Kagaya, R. (1974). A fiberscopic and acoustic study of the Korean stops, affricates and fricatives. Journal of Phonetics 2:161-180.

Kim, C-W. (1965). On the autonomy of the tensity feature in stop classification (with special reference to Korean stops). Word 21 (3):339-359.

Kim, C-W. (1967). Cineradiographic study of Korean stops and a note on "aspiration". Quarterly Progress Report, Research Laboratory, Electronics, M.I.T., No.86.

Kim, C-W. (1970). A theory of aspiration. Phonetica 21:107-116.

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

Müller, E.M. and Brown, W.S. (1980). Variations in the supraglottal air pressure waveform and their articulatory interpretation. In N. Lass, Ed. Speech and Language: Advances in Basic Research and Practice, Vol.4. New York, Academic Press.

Rothenberg, M. (1968). The Breath Stream Dynamics of Simple Released Plosive Production. Basel: S. Karger.

Rothenberg, M. (1973). A new inverse filtering technique for deriving the glottal airflow waveform during voicing. JASA 53:1632-1645.

Smith, B. and Westbury, J. (1975). Temporal control of voicing during occlusion in plosives. Paper presented at 89th meeting of the Acoustical Society of America.

Umeda, H. and Umeda, N. (1965). Acoustical features of Korean "forced" consonants. Gengo Kenkyu 48:23-33.

Westbury, J. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. JASA 73 (4):1322-1336.

Aerodynamic modeling at UCLA

Patricia Keating

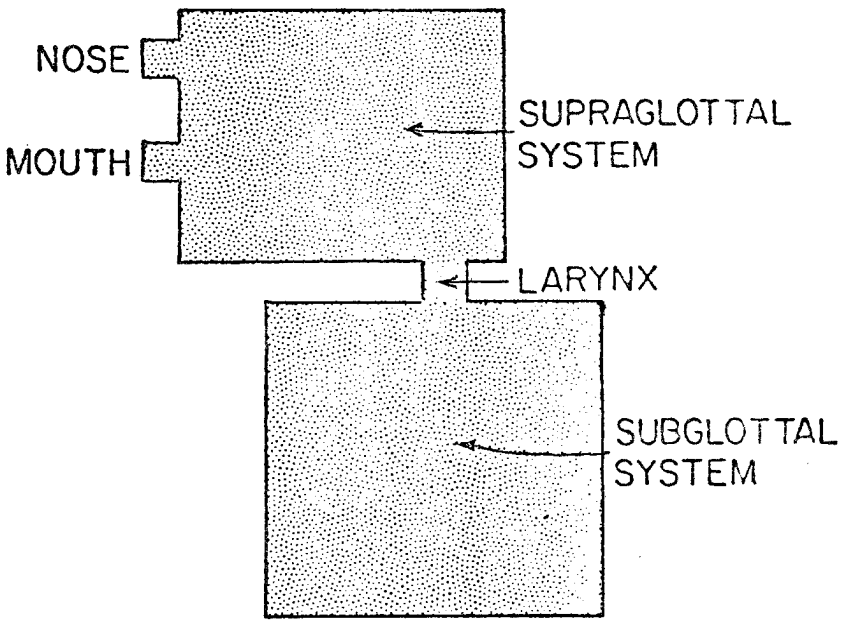
0. Introduction

This report describes the aerodynamic vocal tract simulation currently implemented on the Phonetics Lab's DEC PDP-11/23 computers. The major use of this simulation to date has been in the study of consonants, particularly stop consonant voicing; examples of such work can be seen in other papers in this volume. First I will discuss qualitatively the aspects of speech aerodynamics being modeled, and the kinds of data that such a model will provide. Then I will describe the particular kind of model used in our simulation, outlining some of the basic ideas for phoneticians. Next I will document, in general terms, how the UCLA computer program is used. Finally, I will give an example of work done that will illustrate the use of the model and the kind of results that can be obtained.

1. Speech aerodynamics

Figure 1 shows a schematic of the vocal tract as relevant to a discussion of speech aerodynamics. To a first approximation, the vocal tract consists of two soft-walled cavities, the lungs and the mouth. They are separated from each other by a constriction formed by the vocal cords, and separated from the atmosphere by constrictions at the velum and/or mouth opening. The driving pressure generated by the respiratory system results in airflow from the lungs to the atmosphere via the glottis and one or both of the other openings. Over the course of an utterance, the volumes of both cavities, dimensions of various constrictions, and the mechanical properties of the vocal tract walls may be controlled by a speaker, thereby producing the familiar variations in air pressures and flows which characterize the speech wave.

Figure 1.



The circumstances of air pressures and flows are particularly important in considering vocal cord vibration. A sufficient flow of air through the glottis, from the lungs to the pharynx, is crucial to the occurrence of voicing. According to the myoelastic-aerodynamic theory of phonation (van den Berg 1958), the vocal cords will oscillate when they are suitably adducted and tensed, and when there is a sufficient airflow across them. Such airflow will occur when the pressures above and below the larynx are different: in the usual case, when the pressure in the oral cavity is lower than the pressure in the subglottal system. Air will flow from the region of higher pressure to the region of lower pressure. Calculation of air pressures above and below the larynx, then, will indicate whether voicing could occur for a particular laryngeal state. Since subglottal pressure is relatively constant for most of an utterance, oral pressure is generally the major determinant of that transglottal pressure difference.

Oral pressure depends on how much air is flowing through the glottis into the oral cavity, how much air is flowing through any oral constriction (or the nose) out of the oral cavity, and how much the walls of the oral cavity 'give' in the face of rising oral pressure, counteracting such a rise. Therefore, for example, a larger glottal opening, a smaller oral opening, and stiffer oral cavity walls will all tend to contribute to a higher oral pressure, as contemplation of Figure 1 should make clear. During a vowel, with its large oral opening, oral pressure is essentially the same as atmospheric pressure (taken as a baseline of 0 pressure), so the pressure drop across the larynx (the difference between subglottal and oral pressures) is equal to the subglottal pressure. During a stop, the vocal tract is suddenly and fully occluded, preventing any escape of air and causing pressure behind the occlusion to build rapidly. In this case, then, the pressure drop across the larynx may be small or nonexistent. Consonant release allows the built-up air to escape, and oral pressure drops accordingly, giving a transglottal pressure drop.

2. The analog circuit model

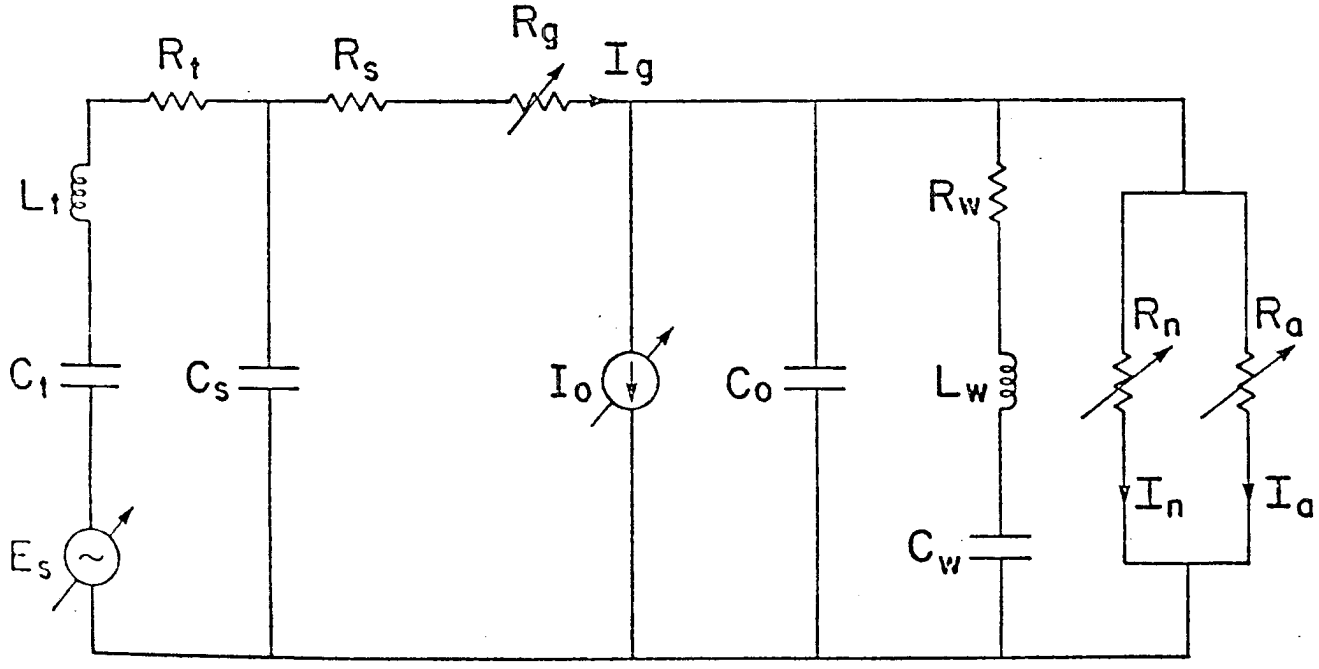
What does it mean to construct a model of speech aerodynamics? In modeling, an idealized representation of a system is constructed to further our understanding of that system. This technique is quite general, and is used in many areas of linguistics besides phonetics. In our case, the simulation is of a physical system, and is numerical, and therefore can be rather precise. One advantage of modeling as a research enterprise is that it forces us to be explicit about our assumptions about the system and about our account of it. Another advantage is that it encourages and focuses the search for new patterns in data.

One way to model speech production would be to build a physical version of a vocal tract, with movable articulators and an air source. This sort of modeling has certainly been tried in the past, but it is less common nowadays than more convenient and more precise numerical simulations of the vocal tract. Such models consist of numerical equations or functions that encode crucial properties of the vocal tract. In some cases these representations are taken directly from the physical system. The aerodynamic model of, for example, Ohala (1976) is of this basic type: airflows and air pressures are calculated directly from things like volume of the vocal tract, and muscular forces.

A less direct kind of model is the electrical analog, which represents the vocal tract as being like a circuit. A circuit is a physical device through which electrical charge can flow; the flow of charge is called current. Various devices or elements may be part of a circuit and will influence this flow in certain known ways. One advantage of a circuit model is that the properties of circuits have been well-studied. If we simply want to know how a circuit would respond to certain inputs, we do not actually need to build it. Rather, we can use circuit theory to devise a set of equations which will describe the behavior of the circuit. Circuit theory also provides mechanical and acoustical correspondances for electrical circuits. Thus a circuit can be designed to represent a non-electrical system, such as a vocal tract; the circuit is then an analog of the vocal tract. The steps in electrical analog modeling are to design a circuit which represents crucial aspects of the vocal tract; formulate the equations describing the behavior of that particular circuit; determine the outputs from those equations for a variety of conditions of interest; interpret those results as phonetic events.

Figure 2 illustrates the circuit used in our work to model vocal tract aerodynamics for low frequency events. This model of the breath-stream control mechanism is derived from work of Rothenberg (1968), who actually built and used a physical circuit. A computer simulation of the circuit model is described by Muller and Brown (1980); essentially the current implementation is described briefly by Westbury (1983). The electrical current moving through the circuit is the analog of volume velocity airflow in the vocal tract. The electrical voltage at any point in the circuit is the analog of air pressure in the vocal tract. The circuit contains five kinds of elements, described below, (plus wire connecting them). These elements, plus their arrangement as a group, represent the crucial aspects of the vocal tract for aerodynamic events.

Figure 2.



The elements labeled E_s and I_o are sources of electrical energy in the circuit: the first introduces voltage, and the second introduces current. I stands for current; the other three I's in the circuit simply label the current coming out of an element. The other three kinds of elements respond to electrical energy, rather than introduce it. All the elements labeled R are resistors, which dissipate energy as heat. The elements labeled C are capacitors, which store energy as electrical energy. The elements labeled L are inductors, which store energy as magnetic energy and in fact introduce magnetic fields into the circuit. Some elements have diagonal arrows through them, which means that their values change over time. Basically, the current starts in the lower left corner, induced by the E_s source, and moves upward through C_t, L_t, and R_t, at which point there is a fork with branches going to both R_s and C_s. Below C_s is a horizontal wire with nothing on it; this is ground or zero where all the charge ends up. Beyond R_s there are several more elements and branches. Every time there is a branching, some current goes one way and some another.

As a reference point, R_g in the top middle represents the resistance to flow presented by the glottis. Basically the glottis is acting like a valve whose opening size can be varied. Everything to the left of R_g represents the subglottal system, and everything to the right of it represents the supraglottal system. The voltage (pressure) source E_s represents the respiratory muscular force, which causes inhalation before an utterance, and counters a drop in subglottal pressure later in an utterance. A set of a capacitor, an inductor, and a resistor in a row, as with C_t, L_t, R_t, and C_w, L_w, R_w, represents the properties of vocal tract walls: in the first case for the trachea and other subglottal cavities, and in the second case for the supraglottal cavity. The capacitor represents the stiffness of the walls; this then includes in the subglottal case what we normally describe as "elastic recoil" and in the supraglottal case, "tenseness". The inductor represents the mass of the walls, and the resistor represents mechanical heat loss inside the walls. The other two capacitors, C_s and C_o, represent the volume of air inside the subglottal and oral cavities, respectively. The volume of the oral cavity is one of the ways in which place of articulation will be reflected. R_s represents other subglottal losses (i.e. it is essentially a fudge factor). The other two resistors, R_n and R_a, are, like R_g, valves: R_n to the nasal cavity and R_a to the atmosphere. If the velum is completely closed so that it totally blocks the flow of air, then it is as if R_n were not in the circuit. Currently, R_n is <u>not</u> represented on our implementation. R_a, representing the oral constriction, is given as three dimensions which depend on place of articulation and vary over time as the constriction is formed or released. The other time varying element, the current source I_o, represents in a single, undifferentiated way, various possibilities for active expansion (or contraction) of the oral cavity, e.g. advancement of the tongue root or jaw movement.

To summarize the elements of the circuit model:

Element	Definition
E_s	Voltage source: respiratory muscle force
C_t	Capacitance (compliance) of tracheal walls; in part depends on surface area of walls
L_t	Inductance (mass) of tracheal walls

R_t Mechanical resistance of tracheal walls

C_s Capacitance (compliance) of air in subglottal system; in part depends on volume of cavity

R_s Other subglottal mechanical resistance

R_g Total glottal resistance; in part depends on size of glottal opening

L_g Reactive component of glottal impedance

I_o Supralaryngeal current source: change in cavity size

C_o Capacitance (compliance) of air in oral cavity; in part depends on volume of cavity

C_w Capacitance (compliance) of oral tract walls; in part depends on surface area of walls

R_w Mechanical resistance of oral tract walls

L_w Inductance (mass) of oral tract walls

R_n Total nasal constriction resistance; no nasal options currently implemented

R_a Total oral constriction resistance; in part depends on size of oral opening

Equally important in using the model are the voltages across the capacitors. Voltage is the electrical analog of pressure: here, either the pressure exerted by air within a volume (voltages across C_s and C_o) or by stretched walls (voltages across C_t and C_w). The two subglottal voltages C_s and C_t are involved in representing elastic recoil. The supraglottal voltage C_o is equal to oral pressure.

The arrangement of the elements in the circuit is dictated by the physical system the circuit is modeling in ways that are well understood by engineers, if not linguists. Given this circuit diagram, an engineer can derive a set of differential equations which describe its behavior. These equations can then be approximated by difference equations which compute changes in pressures and flows for each small unit of time. Such equations are suitable for programming on a small laboratory computer.

A caveat is in order about how voicing is represented in this model. It does not directly represent the vocal cords, so voicing cannot be seen directly as vocal cord vibration. Recall that the conditions on vocal cord vibration include position and tension of the cords, and airflow through them. The position can be fairly well represented via the glottal dimensions. The glottis is modeled as a three-dimensional space, essentially a valve-like opening, which does not vary during voicing. The cross-sectional area of the glottal slit approximates the average glottal area during a vowel, determined over the full duration of a

glottal period. This average area is used in simulations where the glottis is meant to be in a position that would allow voicing, and is used as the final value of any glottal adduction gesture. (The area we use is .04 cm^2; this value is justified by the fact that oral flow during a vowel is about 150 cm^3/sec while the pressure drop across the glottis is about 10 cm aq.) The tension of the vocal cords cannot be represented, since there are no vocal cords. Instead, tension is reflected in our estimation of how much airflow would be required to cause vibration: the tenser the cords, the greater the volume velocity required. We consider this aerodynamic condition in terms of the pressure drop across the larynx. There is some concensus in the literature that a pressure drop of 2 cm aq is necessary to sustain vibration (Lindqvist 1972, Ladefoged 1964, Ishizaka and Matsudaira 1972, Baer 1975), though initiating voicing may require a pressure drop twice as large (Baer 1975). Whenever our model is used to look at voicing, we can only look at pressure across the vocal cords. Typically we assume 'normal' tension of the cords, and the glottal dimensions just described, and determine when the pressure drop is great enough to allow voicing.

Other limitations of this model follow from various simplifications. The treatment of the subglottal system is sketchy. As noted above, various ways of changing the size of the oral cavity are not distinguished. Different parts of the tract are not distinguished, for example, the stiffness of the walls is assumed to be uniform, though obviously this is not the case. Furthermore, the model is valid only for low frequencies (large time intervals).

3. Use of the computer program

The FORTRAN computer program is called VTM, for Vocal Tract Model. Presently it runs on the Phonetics Lab's PDP-11/23 computer under a time-sharing system, without graphic capabilities. However, it produces output files that can be transported to the Phonetics Lab's speech system for displaying and printing.

The following is a complete list of all the input variables under control by the user. Many of them are discussed further below.

User likely to vary from run to run:

CO	volume of oral cavity, with constants
CW, RW, LW	oral wall properties, including area and stiffness
GWID, GLEN	constant glottal dimensions
OWID, OLEN	constant oral dimensions
VCS, VCT	elastic recoil factors in subglottal pressure

User unlikely to vary from run to run:

RHO	density of air
VISCOS	viscosity of air
C	speed of sound
CT, RT, LT	subglottal wall properties, including area and stiffness
CS	volume of subglottal cavity, with constants
RS	other subglottal losses
TINCR	time interval for calculations
NPPE	"sampling rate" for output of calculation results
GTURB	glottal turbulence factor (angle of entry)
OTURB	oral turbulence factor (angle of entry)

Functions over time:

GHEI	distance between vocal cords
OHEI	distance between oral articulators
ES	respiratory muscle force
IO	active change in volume of oral cavity

VTM provides default values, suitable for a medial labial stop, for all of the input variables that are constants. Some of these values are given below. None of them have to be changed in using VTM, but any or all of them may be. In contrast, the functions over time, for ES (respiratory muscle force), GHEI (distance between vocal cords), OHEI (distance between oral articulators), and IO (active expansion of oral cavity), do not have default values. Values at any number of points in time are specified by the user, and VTM linearly interpolates between those values.

Following are descriptions of and values for each of the input variables likely to be changed in VTM.

The variables CO, CW, RW, and COILW together encode the size of the oral tract, the surface area of the tract walls, and the stiffness (or tenseness) of the walls. Values appropriate to typical choices are given on the next page. GLEN represents the dimension of the glottis parallel to the flow of air, i.e., the vertical dimension. The glottal area in the horizontal plane is represented as a rectangle with one fixed and one changing dimension. The fixed dimension, GWID, is the larger of the two dimensions perpendicular to flow. The changing dimension, GHEI, is discussed below. The glottal area is calculated as the product of GWID and GHEI.

OLEN and OWID (and OHEI) are the equivalent oral constriction dimensions; again, OHEI is variable and described below. OLEN is the dimension parallel to flow, and OWID is the larger perpendicular dimension -- for a labial, the width across the lip opening. OHEI is the vertical opening dimension. Again, the area of the constriction is the product of OWID and OHEI. Values have been estimated for John Westbury's vocal tract but are schematic. Good values for OLEN (at the moment before release, from X-rays) are .2 cm for labials, .3 cm for alveolars, and .7 cm for velars. OWID has a small enough effect on outputs that we don't bother to change it.

VCS and VCT do not have such easy physical interpretations; they are variables in the subglottal system representing pressures that together control subglottal pressure via elastic recoil. Roughly, VCS is related to the volume of the subglottal cavity and determines the initial subglottal pressure for a simulation; VCT is related to the surface area of the stretched subglottal walls and determines the change in subglottal pressure, which usually rises when VCT is negative. The default values are for high and very slightly falling subglottal pressure, as for utterance-medial material. Subglottal pressures can be scaled down by varying both VCS and VCT. At the beginning of an utterance, subglottal pressure should rise -- this is done by letting VCS = 0, leaving VCT at default, and varying ES as described below. At the end of an utterance, subglottal pressure should fall -- this is done by leaving VCS and VCT at default, but varying ES as described below. Variation of ES is also described below for changes associated with stress.

The following are possible values for the constant representing the volume of air in the oral cavity, CO, depending on place of articulation:

labial	alveolar	velar
7.16E-5	5.82E-5	4.48E-5

The following are possible values for the constants describing the stiffness of the vocal tract walls, depending on place of articulation. They are ordered with respect to increasing stiffness.

WALLS LIKE LAX CHEEKS

labial	alveolar	velar
RW=6.4	RW=7.27	RW=8.0
COILW=1.68E-02	COILW=1.909E-02	COILW=2.1E-02
CW=1.4973E-03	CW=1.3018E-03	CW=1.1834E-03

DEFAULT CASE -- WALLS LIKE TENSE CHEEKS

labial	alveolar	velar
RW=8.48	RW=9.64	RW=10.6
COILW=1.2E-02	COILW=1.364E-02	COILW=1.5E-02
CW=5.6256E-04	CW=4.9505E-04	CW=4.5E-04

WALLS LIKE NECK WALL

labial	alveolar	velar
RW=18.56	RW=21.09	RW=23.2
COILW=1.92E-02	COILW=2.182E-02	COILW=2.4E-02
CW=2.5458E-04	CW=2.2403E-04	CW=2.0367E-04

There are four variables whose values are functions, not single numbers. These variables, GHEI, OHEI, ES, and IO, do not have default values, and they are all given values at the same time and in the same way. In principle, any variable in VTM can be time-varying, and for a specific application a user may want to do the minor amount of programming required to make some variable a time function, but in general the arrangement described here gives a reasonable balance between accuracy and convenience.

GHEI is the function for changing glottal opening. We represent the mean value over a vibratory cycle as .022 cm. A reasonable value for a spread glottis is .18. For an opening or closing gesture, the user can have VTM interpolate between these. OHEI is the function for changing oral opening. A typical value for an oral opening is .2 or .3; closure for a stop is 0.0. Transition durations for an oral opening or closing gesture must be specified with the OHEI function. 20 msec may be used for a quick trial run, but something like 50-55-70 for labials-alveolars-velars is more natural.

ES represents changes in subglottal pressure through muscular force. When it is zero, subglottal pressure is determined solely by elastic recoil. ES should rise from about -10 cm aq to 0 for initial stops and fall back from 0 to -9 cm aq or so for final stops. A rise or fall like this takes 200 msec in English; we model it linearly. Changes in ES to represent stress look like the desired change in subglottal pressure: e.g. for a rise of about 4 cm aq, ES rises about 4 cm aq with the peak in ES about 20 msec earlier than the desired peak in P_s.

IO represents active expansion or contraction of the volume of the vocal tract, such as jaw lowering, pharynx expansion, or larynx lowering. Its unit corresponds to volume changes (in cc's). 40 cc/sec is a large change typical as a peak value for a [b].

When the RUN command is given, VTM performs all calculations, puts an output file into the LP queue, and asks for new inputs.

Envisioned for the future is a version of VTM that works in terms of physiological, rather than circuit, variables, so that the user of the program need not know anything about circuit elements or about the model to use the program. It may even be desirable to group together variables into such higher-level phonetic inputs as "place of articulation", using default values as sketched above, so that undergraduate students can do simple exercises.

4. An example: voicing in utterance-medial stops

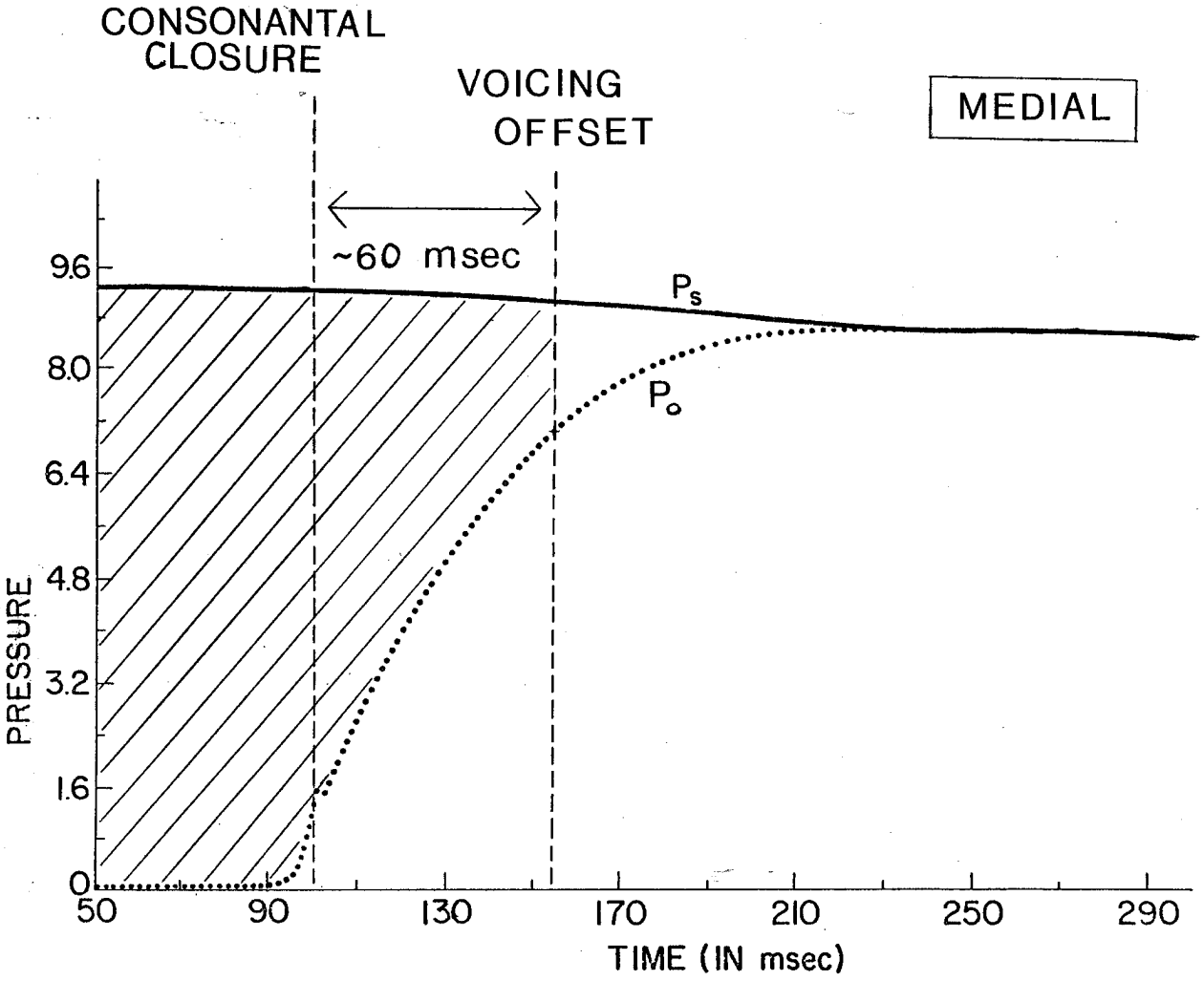
Using the model, it is possible to calculate such things as how P_o will increase in time following the moment of occlusion, as long as sufficient detail about the articulatory states corresponding to its elements is specified. Consider, for example, how P_o will change during a labial stop which occurs utterance medially, between identical vowels. Suppose that all but one of the time-variable elements in the model subject to voluntary control are fixed. Specifically, pressure below the glottis (initially perhaps as much as 10 cm H_2O above atmosphere) derives entirely from elastic recoil of the stretched tissues surrounding the lungs (therefore E_s=0); there are no muscularly induced changes in supraglottal volume (therefore I_o=0), or in the mechanical properties of tissues surrounding the lungs and mouth (therefore RLC_t and RLC_w are constants); and the vocal folds are appropriately adducted and tensed for voicing (glottal area is constant). Only cross-sectional area of the mouth opening (A_a) is allowed to vary, as it must, first to produce a constriction at the lips and then to release it.

Under conditions such as these, pressures above and below the glottis (P_o and P_s, respectively) can be expected to change with time as shown in Figure 3. Note from this figure that the difference between P_s and P_o, though decreasing, is clearly greater than 2 cm aq, the amount thought to be required for voicing maintainence, for the first sixty-odd msec of the 80 msec closure interval. Thus, voicing can be expected during that portion of the intervocalic stop, with offset occurring only late in the closure, within 20 msec of release. The general result, then, is that a labial stop articulation under the aforementioned conditions will naturally be largely voiced.

The relatively lengthy interval of closure voicing for the simulation shown in Figure 3 is 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 -- slows down the decrease in the transglottal pressure drop, and thereby lengthens the interval during closure when voicing is possible. If the walls of the vocal tract were rigid, effective pressure neutralization (and voice offset) would occur within 10 msec of occlusion, as Rothenberg (1968) showed. The value used here for wall stiffness is a more realistic one, and allows voicing to continue for a longer time.

Figure 3.



Of course, a voiceless output could be guaranteed by a glottal spreading gesture during the closure interval. On the other hand, there are several articulatory adjustments which may prolong the voicing interval well beyond the 60 msec or so suggested by Figure 3, producing a fully voiced stop. These include increasing P_s by activating the expiratory muscles; decreasing average area of the glottis and/or tension of the vocal folds; and decreasing P_o by decreasing the level of activity in muscles which underlie the walls of the supraglottal cavity; actively enlarging the volume of that cavity by adjusting positions of the larynx, tongue, and soft palate; or creating a narrow opening between the posterior pharyngeal wall and soft palate (nasal leak). These maneuvers,

occurring singly or in combination, will have their greatest effect on the duration of closure voicing when they occur during the closure interval itself, in concert with the rise in P_o which accompanies vocal tract occlusion. Implementing maneuvers such as these in the model involves specifying how each of the relevant control parameters will vary in time.

Acknowledgements

The implementation of the aerodynamic model described in this report was carried out in the Speech Communication Group at MIT, largely by John Westbury. The results described in Section 4 above are part of research that was a joint effort between me and John, and will be presented more fully in a forthcoming paper. Some of the prose in this report is taken directly from that paper. With support from a grant from NINCDS to Peter Ladefoged, the very primitive computer program used at MIT has been replaced at UCLA with one written by Chris Pettus and Flora Wu, students in our Linguistics & Computer Science major, and me, and documentation has been provided. Thanks are due to the UCLA graduate students who have tried out the program and its documentation.

References

Baer, T. (1975). "Investigation of Phonation Using Excised Larynxes", unpublished doctoral dissertation, MIT.

Ishizaka, K. and M. Matsudaira (1972). "Fluid Mechanical Considerations of Vocal Cord Vibration", Speech Commun. Res. Lab. Mono. No. 8.

Ladefoged, P. (1964). "Comment on 'Evaluation of Methods of Estimating Subglottal Air Pressure'", J. Speech Hear. Res. 7, 291-292.

Muller, E. M., and W. S. Brown, Jr. (1980). "Variations in the Supraglottal Air Pressure Waveform and their Articulatory Interpretation", in Speech and Language: Advances in Basic Research and Practice, Vol.4, edited by N. Lass (Academic Press, New York), pp. 317-389.

Ohala, J. J. (1976). "A model of speech aerodynamics", Report of the Phonology Laboratory (Berkeley) 1, 93-107.

Rothenberg, M. (1968). The Breath-Stream Dynamics of Simple-Released-Plosive Production. Bibl. Phonetica 6.

van den Berg, J. (1958). "Myoelastic-Aerodynamic Theory of Voice Production", J. Speech and Hearing Research, Vol.1, No. 3, 227-244.

Westbury, J. R. (1983). "Enlargement of the supraglottal cavity and its relation to stop consonant voicing", J. Acoust. Soc. Am. 73, 1322-36.

Physiological effects on stop consonant voicing

Patricia A. Keating

Slightly expanded version of
paper presented at ASA meeting, May 1983, Cincinnati

Introduction

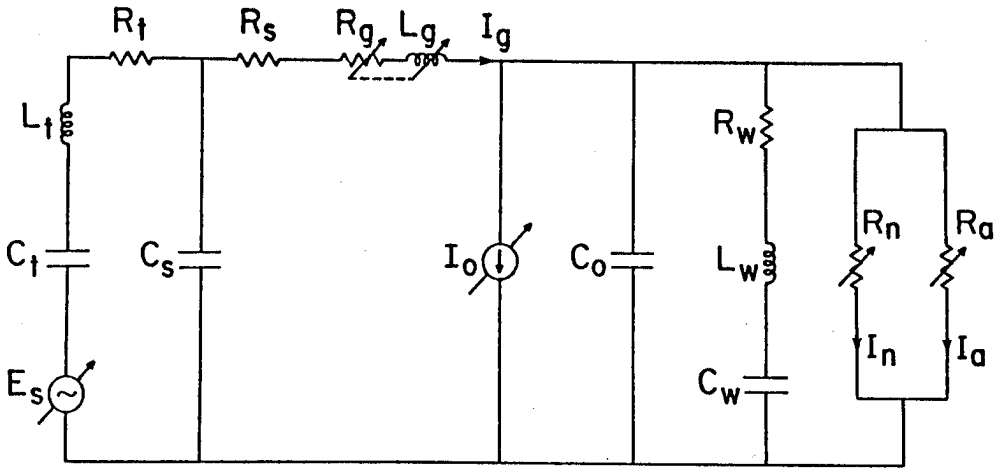
The effect of the voicing distinction on various temporal measures for stop consonants has been much studied. Phonologically voiceless stops are known to have longer Voice Onset Times, that is, more phonetic voicelessness, than phonologically voiced stops. They may also have less closure voicing and a longer total voiceless interval. However, other phonological distinctions besides voicing affect these same timing measures. So, for example, place of articulation is known to affect VOT and closure duration, and stress to affect VOT.

Various researchers, in noting these diverse effects, have offered suggestions that they are due to aerodynamic conditions, as determined by such things as vocal tract volumes, release velocities, and heightened subglottal pressure (e.g. Klatt 1975, Smith and Westbury 1975, Weismer 1980). Although some of these suggestions have been picked up by others and repeated as fact, to date there has been little evidence offered in their support. This paper reports preliminary results of an attempt to explore various possible explanations of temporal effects using an aerodynamic model.

Method

The model, shown in the figure below, is a computer implemented simulation of a circuit analog of the vocal tract derived from Rothenberg (1968) and similar to the model of Muller and Brown (1980). Details about values of circuit elements for this version are given in Westbury (1983). In such a model, voltage is taken as the analog of air pressure, and current as the analog of volume velocity airflow. Values for subglottal pressure, flow through the glottis, oral pressure, and flow through the oral constriction, among other things, are calculated for each small moment in time. Our interest here will focus on when aerodynamic conditions will permit voicing, other conditions being favorable: specifically, when the difference between subglottal pressure and oral pressure is sufficient to permit vocal cord vibration. It will be assumed that a difference of 2000 dynes/cm^2 is necessary to sustain voicing, while a difference of 3000 dynes/cm^2 is necessary to initiate voicing. All results of simulations reported below represent time elapsed until such pressure differences were obtained.

As a source of acoustic data to be modeled, six speakers each of California English, Stockholm Swedish, and Tokyo Japanese were recorded reading real words with stop consonants /b d g p t k/ in initial and medial positions, before nonhigh vowels but after an uncontrolled set of vowels. For English, the 196 words also systematically varied degree of stress; for Swedish, the 372 words also varied pitch accent and stress, and for Japanese the 48 words also varied pitch accent. From a computer oscillographic display, up to three measurements were made for each stop: duration of voiced closure, duration of voiceless closure, and duration of voicing lag. Measurements were averaged across speakers of a language for each word type.



Results

A topic of some interest in phonetics has been the maintenance of voicing during the closures of medial [b], [d], and [g] through passive expansion of the oral cavity, which keeps oral pressure relatively low. The effect of place of articulation on the duration of this medial voicing is illustrated in Table 1, which shows data for [b], [d], and [g] in two English stress conditions and one Swedish pitch-stress condition. The numbers represent msec of voicing after stop closure.

Table 1

	ENGLISH		SWEDISH
	_ main stress	_ 2ary stress	_ main stress
b	55	42	71
d	40	39	57
g	31	24	43

Although such effects used to be thought due to differences in cavity size behind the oral constriction (Smith and Westbury 1975), it is now recognized that the contribution of this parameter is quite small compared to the contribution of the surface area of the cavity walls. Rothenberg (1968), Muller and Brown (1980), and Ohala (1983) all agree that the differences in compliant wall area across place of articulation should produce the observed differences in voicing duration, but none of them have presented supporting data from modeling. This is the goal of our first modeling exercise. The three places of articulation were simulated at three degrees of wall impedance following Ishizaka et al. (1975). No subglottal or glottal differences were assumed. Differences in place of articulation can be represented as differences in cavity volume, surface area of walls as it influences impedance parameters, the dimensions of the oral constrictions, and the speed of the closing gesture. However, independent manipulations of these variables indicate that the largest effect by far on closure voicing is due to the wall surface variables. Table 2 gives the duration

of voicing to be expected from simulations for [b], [d], and [g] when the walls are like lax cheeks, when they are like moderately tensed cheeks, and when they are like the neck walls. The linguistic data in Table 1, with which the simulations in Table 2 can be compared, suggest wall values that were slightly more tensed than the tense cheek values used here. But the point is that for any given setting, place of articulation differences will produce acoustic differences in voicing maintenance: fronter places, with greater surface area, have more voicing.

Table 2

	lax cheeks	tense cheeks	neck
b	145	54	25
d	124	46	21
g	105	36	13

[b], [d], and [g] are not the only stops to have voicing during closure: for most speakers, especially of English, [p], [t], and [k] will have at least one or two pitch periods of closure voicing. Table 3 shows durations of such closure voicing in English before reduced vowels, in Japanese between High and Low pitches, and in Swedish in Accent 1 words between reduced and stressed vowels. As Rothenberg suggested, the vocal cords' opening gesture for voiceless segments will allow some vibration, presumably breathy, before their separation makes voicing impossible. Will wall surface area differences across place of articulation produce the voicing duration differences observed?

Table 3

	ENGLISH	JAPANESE	SWEDISH
	_ reduced V	H _ L	reduced V _ stressed V
p	13	10	16
t	7	10	12
k	6	5	8

The first column in Table 4 below shows (as closure voicing in msec) the result of simulations in which all three stops have a constant glottal gesture beginning at closure and walls like tense cheeks. These simulations indicate that the differences due to wall surface area are quite small, with only a 3 msec difference between labial and velar. The differences in the acoustic data in Table 3, while also small, are larger, and may indicate some additional mechanism. One such mechanism is suggested by observations on the timing of the glottal gesture relative to consonant closure and to consonant release. Lofqvist (1980) and others (e.g. Lofqvist and Yoshioka 1981) have noted that for aspirated stops the glottal gesture appears to be timed to begin at the moment of stop closure and to reach its peak opening value within 20 ms before the moment of release. That is, the time to peak glottal area is proportional to closure duration. Extending this observation to the case at hand, we can note that stops at different places of articulation differ in closure duration: fronter places

have longer closures. As long as peak glottal area does not vary across place of articulation, then fronter places of articulation will have more time for the vocal cords to travel the same distance, that is, slower glottal gestures. With these assumptions, the right order of magnitude of voiced closure will result; the second column in Table 4 shows the durations of closure voicing in msec that result when the labial closure duration (and therefore the time to peak glottal area) is 10 msec longer than the alveolar, and 20 msec longer than the velar.

Table 4

	wall effects only	add varying glottal gesture
p	15	15
t	13	12
k	12	9

Consider next the effect of stress on medial [b d g] voicing. First, English [b d g] before reduced vowels are typically voiced throughout their closures, which are quite short. But if there is a break in voicing, as is common for stops before non-reduced vowels, then there is more stop closure voicing before a more stressed vowel. Table 5 shows durations of closure voicing for Swedish and English [b d g] as affected by the stress on a following vowel.

Table 5

	SWEDISH		ENGLISH	
	_ main stress	_ other V	_ main stress	_ other V
b	71	56	55	45
d	57	31	40	34
g	43	39	31	28

Main stress increases the duration of voicing. To see why this rather unexpected result should hold, consider how stress is effected through increased respiratory muscle activity, resulting in greater subglottal pressure. Data of Ladefoged and colleagues (Ladefoged 1967) shows that there is muscle activity before a stressed vowel, that is, during the closure of the preceding consonant. These data, and similar data of Lieberman (1967), show an increase in subglottal pressure of from 1 to 5 cm aq, with peak pressure after the onset of the vowel, and relatively smooth increases and decreases of about 100 msec around that peak. Simulations indicate that a peak in subglottal pressure will lag a respiratory force peak by about 20 msec. Table 6 compares the effect (msec closure voicing) on labials of such a muscularly-induced boost in subglottal pressure when it begins at the moment of closure and lasts longer, and when it begins half-way through closure and is somewhat shorter, with the ordinary case we saw before for b. Under the two boost conditions, both subglottal and oral pressure will be higher during closure, but the subglottal pressure will be proportionately higher, such that closure voicing will be increased in duration by about 5 msec, on one simulation, or will last longer than the consonant, on the other.

Table 6

Long P_s boost	Short P_s boost	No P_s boost
100+	59	54

The finding that stress on a following vowel favors voicing, all things being equal, does further work for us. Table 7 compares lag VOT in initial stops in English before main stressed, secondary stress, and reduced vowels. Stress on the vowel following [b], [d], and [g] also decreases the VOT value, as has been noted by Lisker and Abramson (1967).

Table 7

	_ main stress V	_ 2ary stress V	_ reduced V
b	10	10	13
d	13	16	18
g	24	27	27

An initial /b/ was simulated under two conditions, with and without the extra muscularly-induced boost in subglottal pressure. In the case without, VOT is about 8 msec, but if we give an extra respiratory push, subglottal pressure will be relatively higher than the oral pressure, and voicing begins at 6 msec, that is, 2 msec earlier. It's not clear that the magnitude of the simulated effect is sufficient compared to the data, but the direction of the effect found is encouraging.

Conclusions

Although more work clearly remains to be done, we have seen that certain observations about acoustic effects of place of articulation and stress may be accounted for as consequences of other physiological variables. The effects of place and stress on the duration of closure voicing for [b d g] may derive from independently necessary factors, such as wall area and respiratory force. The effect of place on [p t k] may involve a more arbitrary factor, the starting time and speed of the glottal gesture. Further work, with more attention to small cross-language differences, may motivate these interarticulator timing differences. At the same time, it may elucidate how the physical properties of the speech production system constrain acoustic variation.

Acknowledgement

This research was supported by a grant from NIH to Peter Ladefoged.

References

Ishizaka, K., J. C. French, and J. L. Flanagan (1975). "Direct Determination of Vocal Tract Wall Impedance", IEEE Trans. Acoust., Speech Signal Process. ASSP-23 (4), 370-73.

Klatt, D. H. (1975). "Voice-onset time, frication, and aspiration in word-initial consonant clusters", J. Sp. Hear. Res. 18, 686-706.

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

Lieberman, P. (1967). Intonation, Perception, and Language (MIT Press, Cambridge).

Lisker, L., and A. S. Abramson (1964). "A cross-language study of voicing in initial stops: acoustical measurement", Word 20, 384-422.

Lisker, L., and A. S. Abramson (1967). "Some effects of context on voice onset time in English stops", Lg. Speech 10, 1-28.

Lofqvist, A. (1980). "Interarticulator programming in stop production", J. Phon. 8, 475-490.

Lofqvist, A., and H. Yoshioka (1981). "Interarticulator programming in obstruent production", Phonetica 38, 21-34.

Muller, E. M., and W. S. Brown, Jr. (1980). "Variations in the Supraglottal Air Pressure Waveform and their Articulatory Interpretation", in Speech and Language: Advances in Basic Research and Practice, Vol. 4, edited by N. Lass (Academic Press, New York), pp. 317-389.

Ohala, J. J. (1983). "The Origin of Sound Patterns in Vocal Tract Constraints", in The Production of Speech, edited by P. F. MacNeilage (Springer-Verlag, New York), pp. 189-216.

Rothenberg, M. (1968). The Breath-Stream Dynamics of Simple-Released-Plosive Production, Bibl. Phonetica 6.

Smith, B. L., and J. R. Westbury (1975). "Temporal control of voicing during occlusion in plosives", paper presented at April ASA meeting.

Weismer, G. (1980). "Control of the voicing distinction for intervocalic stops and fricatives: some data and theoretical considerations", J. Phon. 8, 427-38.

Westbury, J. R. (1983). "Enlargement of the supraglottal cavity and its relation to stop consonant voicing", J. Acoust. Soc. Am. 73, 1322-36.

Universal phonetics and the organization of grammars

Patricia A. Keating

Chapter to be published in a book edited by V. A. Fromkin

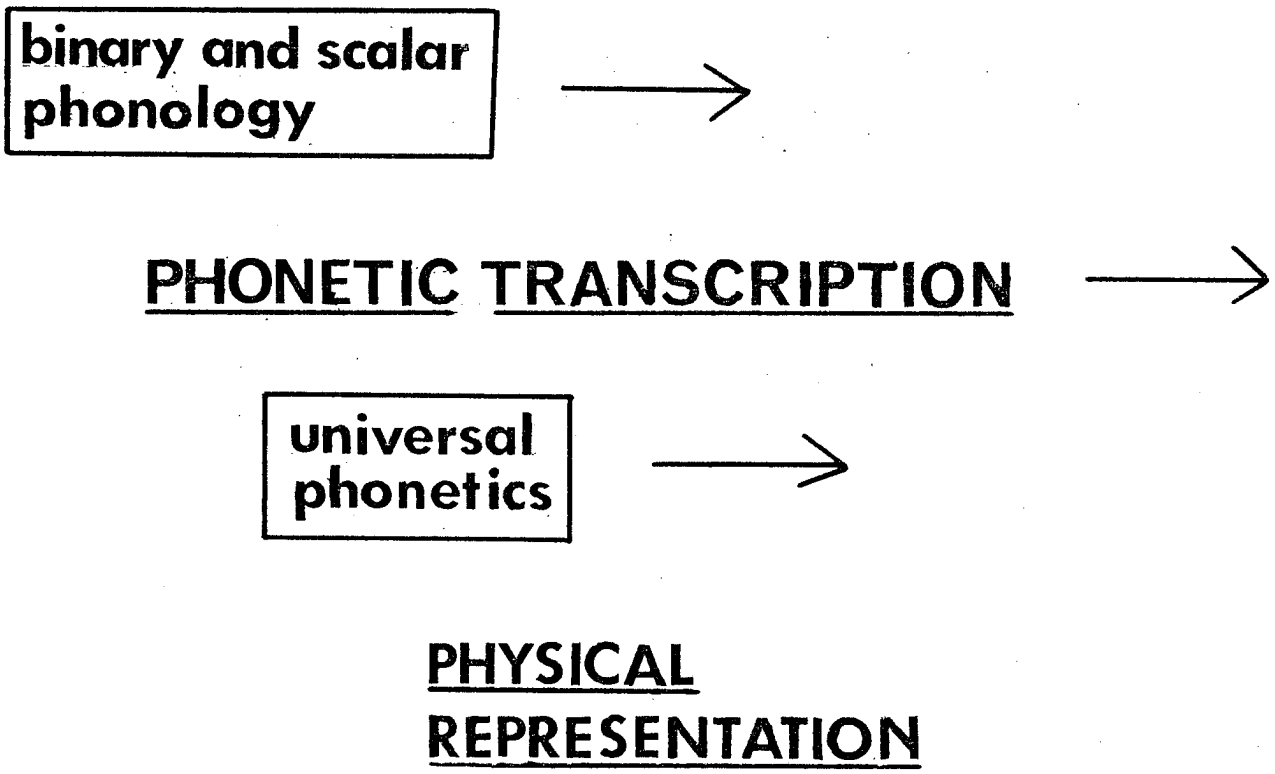
Introduction

Phoneticians have long been interested in the relation between phonetics and phonology, and especially so since the rather explicit proposals of Chomsky and Halle (1968). Much of the attention has focused on the nature and substance of the phonetic feature system; Ladefoged (e.g. Ladefoged 1971, Ladefoged 1980) has been a notable participant in this discussion. However, it is of some theoretical interest that in the SPE model, phenomena that might be called 'phonetics' are found in two separate places. On the one hand, the phonetic rules that convert binary into scalar feature values are part of the phonological component of the grammar. On the other hand, the other part of phonetics, the part actually called phonetics, is not technically in the grammar. It is a largely universal and predictable component which translates a segmental phonetic transcription into continuous physical parameters. Broadly speaking, this extra-grammatical physical phonetics is the locus of many of the traditional (as well as recent) concerns of phoneticians -- articulation, timing, coarticulation, etc. In this paper I would like to consider the division of labor between phonology and phonetics in more detail, and suggest a direction for revision in the model.

The figure below gives a schematic view of the relevant parts of the SPE model. First, the phonological component of the grammar contains both phonological rules that operate on binary-valued features, and language-specific phonetic detail rules. The phonetic detail rules convert binary phonological feature specifications into quantitative phonetic values, called the 'phonetic transcription', or systematic phonetic representation. These rules in part depend on universal phonetic constraints concerning possible combinations and contrasts. "Given the surface structure of a sentence, the phonological rules of the language interact with certain universal phonetic constraints to derive all grammatically determined facts about the production and perception of this sentence. These facts are embodied in the 'phonetic transcription'." (p. 293) According to Chomsky (1964), phonological rules apply until a representation in a universal phonetic alphabet results. The phonetic transcription represents "what the speaker of a language takes to be the phonetic properties of an utterance, given his hypotheses as to its surface structure and his knowledge of the rules of the phonological component" (SPE, p. 294). It is not "a direct record of the speech signal", and is only one "parameter determining the actual acoustic shape of the tokens of the sentence". The physical utterance itself is not generated by the grammar; the phonetic transcription is the terminal output of the grammar.

As such, the phonetic transcription must be further interpreted (translated, spelled out, realized) as a physical phonetic representation by a phonetic component which is not technically part of the grammar. The assumption here is that, with the right phonetic representation, any utterance in any language can be interpreted by a set of phonetic conventions. The translation from discrete segments to articulations that exist in time is treated as being automatic; the phonetic component includes, for example, "the different articulatory gestures

and various coarticulation effects -- the transition between a vowel and an adjacent consonant, the adjustments in the vocal tract shape made in anticipation of subsequent motions, etc." (p. 295).[1] Chomsky and Halle further suppose that these phonetic conventions are universal rules -- that the same phonetic rules can interpret a phonetic transcription in any language. Although not strictly necessary within the general model, certainly this view is an appealing one. The phonology contains the language-specific statements required to produce a detailed enough transcription to allow phonetic interpretation, and the phonetics converts that transcription into a physical utterance in a quite automatic way. This distinction between language-specific rules vs. automatic low-level phonetic rules is in some ways similar to the distinction between "extrinsic" vs. "intrinsic" allophones (MacNeilage 1970), or between "soft" and "hard" coarticulation (Fujimura and Lovins 1978). Though apparently not by Chomsky and Halle, the phonetic rules are often thought to be directly motivated by, or be identical to, physical constraints on articulation or perception.



The SPE model of grammar thus specifies a very constrained relation between phonological and phonetic representations. The use of phonetic features in phonological representations ensures that lexical representations and phonological rules can be evaluated for their phonetic naturalness (the Naturalness Condition, Postal 1968). However, the phonetic representation may, under this theory, be much broader or much narrower than a traditional transcription. Here the phonetic transcription is defined by its position in the model between the phonological and phonetic components. It follows, then, that the fewer universals of phonetic realization are posited, the narrower the phonetic transcription will have to be, while the more such phonetic universals

there are, the broader the transcription will be. Suppose, for example, that speakers interpreted all phonetic properties as grammatical ones. In that case, the phonetic transcription as defined by the theory and generated by the grammar would be much narrower than the traditional segmental phonetic transcription. For example, such a view is found in Pierrehumbert (1980)'s work on intonation. The language-specific quantitative rules in her system directly output something very close to a physical utterance.

How could a more traditional phonetic transcription be maintained if speakers interpreted phonetic properties as grammatical? Clearly we would have to say that the interpretive rules of the phonetic component are part of the grammar. This move would recast the phonetic component so that it is no longer mainly the domain of universal conventions. Would such a revision be completely arbitrary, just to preserve the phonetic transcription? We might propose that, rather than provide automatic aspects of interpretation, the phonetic component derives any aspect of a representation in which continuous time is involved (cf. Anderson 1974). Then the phonetic component could still do much of the work of deriving the physical phonetic form of an utterance, and the transcription could be a fairly broad one.

Some revision in the SPE model is required in one of these directions simply because in fact there does not appear to be a well-defined body of phonetic universals that operate automatically across languages. Phoneticians are aware that many supposed universal rules of phonetic interpretation have exceptions. What is to be made of these exceptions, of the phonetic rules, and of the phonetic component of the SPE model? Is there any role for a universal phonetic component, in or out of the grammar? And what is the relation of near-universals to physical phonetic constraints?

Questions about the nature of phonetic rules do not disappear if we reject the SPE model of grammar. In fact, the importance of the issue is only increased when we look at alternative theories proposed in the seventies. Some rejections of the SPE model have been based on phonetic naturalness as a defining property of phonology. In these views, naturalness, in turn, is generally linked to the mechanisms of speech production and to phonetic universals (e.g., Natural Generative Phonology (Hooper 1976)). There, phonological rules are constrained to be those natural rules that are exceptionless because they are directly physiologically motivated.

In this paper I will examine three known phonetic patterns in light of the SPE model and the discussion above. With the first case, I will discuss the fact phonetic patterns are not necessarily automatic results of speech physiology. With the second case, I will illustrate that they need not be universal, and that they can operate as abstract phonological rules. With the third case, I will consider the limitations of physiology in determining phonetic patterns. While none of these types of observations are original, taken together they will lead to some tentative proposals about the place of phonetic patterns in grammars.

Intrinsic vowel duration

Let us first consider the case of intrinsic vowel duration. In most if not all languages low vowels such as [a] and [æ] are longer than high vowels such as [i] and [u], all things being equal (Lehiste 1970). Not only can this phonetic pattern be observed across languages, but a physical explanation has been

suggested. As Lehiste notes, lower vowels require a greater articulatory movement; if movement velocity is nearly the same across vowels, lower vowels will be longer. Furthermore, the observed differences in vowel duration could be accounted for by automatic biomechanical effects rather than by deliberate temporal control. Lindblom (1967) provided an explicit account of such an automatic effect with a mechanical model of jaw activity. Of course, since Lindblom mentioned that compensations in vowel durations could be made, he did not intend the model to account automatically for all aspects of intrinsic vowel duration. However, his work is important because in principle it could provide such an account. Lindblom's model showed that if the force input to jaw lowering muscles had the same duration but different amplitudes for different vowels, biomechanical sluggishness would automatically result in the correct vowel duration pattern. That is, if the jaw gets a harder, but not longer, send-off for lower vowels, which translates automatically into longer movements, then no explicit timing representation is needed. By hypothesis, all vowel heights have the same representation for duration at every point in their production. Thus no information about intrinsic vowel duration need be included in a grammar, since intrinsic vowel duration patterns can be accounted for automatically in a universal component. Such a view is compatible with what Fowler (1980) called 'extrinsic timing' models, in which time is never specifically included in the plan for an utterance, but is introduced only in production. That is, here is a case where, apart from any cross-linguistic data, modeling of the speech production mechanism supports an automatic phonetic universal.

An experiment by Westbury and Keating (1980) investigated this claim about speech production in a physiological study of spoken vowels. Electromyographic techniques were used to study the force input to a jaw lowering muscle for vowels, the anterior belly of the digastric, or ABD. Three American speakers read items of the form /sVts/ 15 times each, where V = each of ten English vowels. We recorded simultaneously on channels of FM tape the speech signal from a microphone, the mandible displacement from a strain gauge device attached to a tooth splint, and the EMG signal from the ABD as measured with hooked wire electrodes. We then measured the acoustic vowel duration from the speech signal, the extent and timing of jaw displacement from the movement signal, the EMG duration from the EMG waveform, and the EMG maximum amplitude from the rms time envelope.

Our results replicated the earlier finding by others of intrinsic vowel duration: lower vowels are longer in acoustic duration than higher vowels, especially (but not crucially) if the English phonological distinction between tense/lax or long/short vowels is taken into account. The measurements also showed (as expected) that lower vowels have a lower jaw position than higher vowels. In addition, the two measures were statistically correlated: the vowels with the lower jaw position had longer acoustic durations. The longer durations were due to longer travel times, not longer steady states. Thus we obtained the data enabling us to address the question of force input. We found that the EMG duration and maximum amplitude both showed the same pattern across vowels, with low vowels having longer durations and higher amplitudes of EMG activity, correlating with jaw displacement. That is, more extensive and longer movements are made with a force input that is both longer in duration and higher in amplitude: the ABD muscle fires longer and more actively; loosely speaking, it pushes both longer and harder to go farther.

We conclude, then, that at the level of neural control tapped by measuring EMG activity, vowels <u>are</u> represented as having different durations, since the

muscle firing varies in duration, like the acoustic vowel signal. Thus vowel duration differences are not due directly to sluggishness of the jaw; rather, they are controlled as such. This result in itself does not show that vowel duration differences must be language-specific, or represented in the grammar: durations could be provided at a very late stage in production before motor commands are issued, by the phonetic component. However, if vowel duration is a controllable parameter, it is in principle available for language-specific manipulation. Thus we should expect to find languages with different vowel duration patterns, just as we found languages without the expected vowel shortening. Such patterns could be seen, for example, when low vowels are considered phonologically short, and are produced with short travel times and high velocities.

We may still wonder why so many languages have similar patterns of intrinsic vowel durations. Though the pattern is not a necessary one, it must be convenient in some sense. It may be that some physical patterns and movements are preferred over others because of general principles of economy of effort and motor control; see, for example, Nelson (1980). The point here, though, is that such principles must be more subtle than absolute mechanical constraints of the sort that might have been proposed. The physical factors clearly influence vowel duration, but they do not control it.

Extrinsic vowel duration

Consider next the general finding that vowels are shorter before voiceless obstruents than before voiced obstruents or sonorants[2]. Chen (1970) surveyed a number of languages, including some described in the literature, and found such vowel duration differences in all of them. Of seven languages studied, all showed at least a 10% difference in vowel duration. This was so whether the vowel and consonant were word-final and tautosyllabic, or whether the vowel and consonant were word-medial and heterosyllabic. Chen suggested that some contextual durational difference is universal and physiologically determined, although languages may individually exaggerate this difference by rule, e.g. English. Although there were problems with Chen's cross-language comparisons[3], it has generally been accepted that vowel length differences depending on consonant voicing constitute a phonetic universal, albeit one whose mechanism is not understood. Fromkin (1977) uses this result to argue that the vowel duration effect is given by phonological rule in English (that is, is represented in the phonetic transcription), but is automatically supplied by universal phonetic conventions in other languages without the exaggeration. However, the pattern is not a universal one, and it must be given by rule even in some languages that do not exaggerate the effect.

As part of a study on Polish voicing contrasts (Keating 1979), Polish vowel durations before voiced and voiceless consonants were measured for the pair rata - rada. Polish, like other Slavic languages, has a rule of word-final devoicing, so there are no voicing contrasts at the end of isolated words. Thus the durational phenomenon can only be studied in medial position, though based on Chen's survey and on the English studies cited below a robust difference is still to be expected. Twenty-four speakers in Wrocław, Poland, were recorded reading this pair, and durations of the stressed syllabic nuclei were measured from a computer implemented oscillographic display at the Brown University Phonetics Lab. The mean duration of [a] before [t] was 167.4 msec, and of [a] before [d], 169.5 msec. The ratio of these two means is .99. In addition, the ratio of the

two vowel durations was computed for each individual speaker. The mean of these 24 ratios is 1.0. These data indicate that Polish vowel duration does not vary systematically according to the voicing of the following consonant.

Comparable data for English vowels before medial stops has been collected by Sharf (1962) and Klatt (1973). The pre-voiceless/pre-voiced ratios they obtained were .75 and .79, respectively. A higher ratio, .89, was obtained by Port (1977), using sentence contexts rather than word lists. (Since English flaps its medial alveolar stops before stressless vowels, these data are for vowels before labials and velars.)

The finding that Polish, unlike English, does not shorten vowels before voiceless consonants was extended by recording speakers of Czech. As both Czech and Polish are West Slavic languages they are similar in many ways, but Czech has phonemic vowel length contrasts. Thus it seemed possible that Czech would also fail to differentiate vowel durations according to consonant voicing, so that vowel duration could be reserved for the phonemic length contrast. Three native speakers of Czech read several words of the following form:

$$\begin{Bmatrix} p \\ ml \end{Bmatrix} \begin{Bmatrix} a \\ a: \end{Bmatrix} \begin{Bmatrix} t \\ d \end{Bmatrix} \quad V \qquad (C)$$

The number of phonemic short and long vowels was balanced. The mean duration of vowels before [t] was 193.7 msec; before [d], 204.2 msec. The ratio of these two means is .95, and the mean of the individual ratios is .98. Thus there is a slight tendency for vowels to be shortened before voiceless consonants, but the difference in durations did not reach statistical significance (t_{30}=-.37, $p>.20$). In sum, neither Czech nor Polish disyllables show the supposed universal vowel shortening before voiceless consonants.

One line of explanation that has been offered for vowel length differences involves the fact that closure interval durations also vary with voicing, and are inversely related to the vowel durations. That is, voiceless stops have longer closure intervals than do voiced stops. In English and presumably other languages with vowel lengthening, the two ratios, vowel and closure, essentially balance each other, so that the syllable duration is relatively constant. What happens to closure, and syllable, durations in Polish? Closure durations for the same 24 pairs were also measured. The mean duration for [t] was 130.1 msec, and for [d], 91.5 msec. The ratio of these means is 1.42, and the difference is statistically significant (t_{23}=8.81, $p<.001$). Comparable data for English labial stops is Lisker (1957)'s ratio of 1.60, and Port (1977)'s ratio of 1.35 (again, from sentence contexts). Thus Polish, like English, has longer closure durations for voiceless stops. Because Polish shows the closure but not the vowel effect, its syllable durations are not balanced.

This finding indicates that the vowel shortening effect, in those languages where it occurs, is not physiologically determined by the closure duration effect. Of course, there could still be some non-physiological relation between closure and vowel duration that some languages could choose to implement. For example, language-specific prosodic factors like stress or rhythm could make it desirable to balance intrinsic syllable durations. This factor may operate more powerfully in a language like English, with variable stress and vowel reduction, than in a language like Polish, with fixed stress.

Thus the possibility that vowel shortening before voiceless consonants is an (automatic) phonetic universal is not supported by an investigation of Polish and Czech. Further counterevidence from Saudi Arabic is found in Flege (1979). He found that long /a:/ was not significantly longer before word-final /d/ than /t/. Therefore, we know that this rule cannot be placed in a universal phonetic component because it does not occur universally across languages. Rules of phonetic vowel duration as a function of a following consonant's voicing must be language-specific.

Furthermore, Chen's study shows that exceptions to the phonetic pattern take still another form. For example, Chen found a vowel duration difference in Russian completely comparable to that in other languages, although a footnote indicates that all the final consonants determining the vowel durations were voiceless, Russian having a rule of final devoicing[4]. The duration pattern was apparently determined by underlying values of the voicing feature. In the same way, vowel duration for speakers of some English dialects varies before voiced flaps according to underlying stop voicing values (Fox and Terbeek 1977). Clearly, if vowel durations can be determined by underlying phonological values for voicing, then the relation between vowel duration and voicing cannot be automatic and physiological. It is important to realize that these cases are actually counterexamples to the pattern at the systematic phonetic level, since in Russian longer vowels occur before voiceless consonants, and in the English dialects shorter vowels occur before voiced consonants. The pattern is clear only at some point in the derivation before the phonetic transcription.

At the same time, there is obviously a trend across languages and across phonological rules that must be accounted for. We can summarize the possibilities as follows: languages can show no vowel durational differences, or they can show some kind of differences which relate shorter vowels to following voiceless obstruents. If they do show such a pattern, they can do so at either the phonetic or phonological level. No language shows durational effects in which vowels are shortened before all voiced consonants and lengthened before all voiceless consonants. It is as if there is a possible patterning available to languages: vowels may be shorter before voiceless consonants. The reverse pattern is not available in this way. Thus we find languages like Polish and Czech, with no difference, languages like French and some English dialects, with shorter vowels before phonetically voiceless consonants, and languages like Russian and German, with a phonologically conditioned pattern.

This example of extrinsic vowel duration patterning shows that a supposed phonetic universal is not in fact universally attested. Because of this fact, and because the extent and level of duration differences varies across those languages with the pattern, the pattern cannot be automatic or predictable. Each language must specify its own phonetic facts by rule. Possibly, following Fromkin 1977, we could say that languages with an exaggerated pattern, and languages with no pattern, must include a rule in their grammars. In addition, languages whose patterns are not exaggerated but operate on phonological representations must also include a rule in their grammars. This leaves languages with an unexaggerated duration difference that is entirely phonetically conditioned: following Fromkin, this pattern could be provided by the phonetic component. Obviously, however, such a treatment entails a change in the conception of the phonetic component. Rather than a phonetic universal that is predictable and automatic, that phonetic statement would represent one special case, simply a kind of "elsewhere" condition on phonetic detail. Alternatively, the phonetically conditioned cases could be treated exactly like the phonologically conditioned cases, by a grammatical rule.

As in the intrinsic duration case, it appears that the role of the phonetics is to provide a pattern that might be prefered. Within any one language, however, vowel duration is controlled by the grammar, even though it is a low-level phonetic phenomenon. While it is a good idea to continue looking for phonetic universals that would support a model of automatic phonetic interpretation, it seems more likely that our eventual model will incorporate phonetic rules of timing into the phonology.

Voicing Timing

In the case of the occurrence and timing of stop consonant voicing, each of the investigative methods considered above has been employed by a number of people. Cross-language surveys have revealed patterns that must be explained, and modeling studies have tried to provide some explanations. What is interesting is that none of the patterns found are universal, yet they are good examples of phonetic "naturalness". Thus these patterns are a key to the relation between physical motivations, phonetic rules, and the grammar.

The sort of patterns I have in mind are exemplified as follows. Surveys of phoneme inventories (e.g. Maddieson 1983) produce two major observations. First, voiceless stops are generally preferred to voiced stops, especially for geminates. Second, the extent of this stop consonant preference re voicing varies according to place of articulation, with further front stops being more likely to be voiced. Thus some languages have /b/ but no /p/ (labials favor voicing), or /k/ but no /g/ (velars favor voicelessness). Surveys of allophone occurrence and detail lead to similar conclusions. In most environments, voiceless unaspirated stops are favored -- even in intervocalic position, contrary to popular belief (Houlihan 1982; Keating, Linker, and Huffman 1983). Place of articulation effects on the duration of voicing and of aspiration can be observed across languages (Lisker and Abramson 1964), although various exceptions have been noted. In this section I will confine discussion to the more categorial effects on voicing discussed in Keating et al. (1983), namely, the position-in-utterance preferences seen in unrelated languages.

The best-known work on physiological motivations for voicing patterns in general is probably that of Ohala (much of it summarized in Ohala 1983). Ohala has used a simple model of breath-stream dynamics to illustrate the common observation that voicing requires glottal airflow, while stop occlusion impedes such airflow, and in this sense stop occlusion and voicing are at odds with each other. Thus it is understandable that voiceless stops should be more common than voiced. He also used the model to reason about the further patterns found. Drawing on other modeling work by Rothenberg (1968) and Müller and Brown (1980), Ohala stressed the role of passive and active expansion of the vocal tract walls in allowing airflow, and hence voicing, to continue during stop occlusion. Wall expansion is related to findings about place of articulation in that the further front the occlusion, the more expandable vocal tract wall area there is between glottis and occlusion. Thus we should expect further front places to allow voicing continuation more easily than further back places.

Westbury and I, together and separately, have looked in more detail at effects of place of articulation, of position in utterance, and of stress on Voice Onset Time (VOT) and on closure voicing duration. A model of voicing based on Rothenberg's was devised, and is described in more detail in Westbury (1983).

It allows us to vary over time the subglottal pressure, the position of the vocal cords, the oral constriction in three dimensions, and the stiffness of the vocal tract walls. We use results from an X-ray study (Westbury 1979) and a tracheal puncture study (University of Texas Phonetics Lab unpublished data) for constriction and pressure data, and other, published, data such as glottal opening, as inputs (see Westbury 1983 for references). The computer program takes these inputs, and calculates the resulting airflows and air pressures in the vocal tract. From the airflow through the larynx we can see exactly when voicing should occur. In the case of position-in-utterance effects, our results (Westbury and Keating 1984) were clear. We compared initial, intersonorant, and final positions, assuming that the only difference across them was the subglottal pressure being generated. We assumed that the vocal cords were equally ready to vibrate in all three positions, and that the oral gestures' closures and velocities were the same, except (non-crucially) that initial closures were longer. Such modeling showed that the pressure differences result in three different acoustic patterns. In initial position, voicing does not occur until after consonant release with these inputs; in medial position, voicing continues from the preceding sonorant through most but not all of the stop occlusion; in final position, voicing continues into the beginning of the occlusion but ceases earlier than in medial position.

A preference of languages for voiceless unaspirated initial and final allophones is thus seen to arise from the physical operation of the speaking device. What does this preference explain? It may be useful to compare our account of final stop voicelessness with Dinnsen (1980)'s discussion of supposed aerodynamic explanations of phonological rules of final devoicing. He distinguishes explaining a rule's structural description (here, that it affects final stops) from explaining its structural change (here, that it devoices them) and from explaining why there should be any rule in the first place (here, some difficulty posed by final voiced stops); he says that only the structural description is explained by e.g. Ohala. Our explanation is different from Ohala's[5], however, and goes further in illuminating a final devoicing rule's structural change and arguably its motivation. This improvement comes from carefully quantifying the articulatory conditions that hold before the stop consonant, the acoustic characteristics of a stop in which those conditions are changed only minimally, and the acoustic characteristics of stops in which those conditions are changed more drastically. A devoicing rule specifies a structural change most in accord with the result of a minimal change in articulatory conditions.

However, as Dinnsen emphasizes, a phonological rule exists independently of such a phonetic motivation. That this must be so in the case of final devoicing is shown by the fact that our motivation applies only to position in utterance effects. Position in utterance is not the same as position in word, and many linguistic rules and constraints operate in the word domain. For example, instances of word-final devoicing in utterance-medial position are phonetic counterexamples to the patterns generated by the model. They may serve to demarcate word boundaries in running speech, for example, but are no longer directly physically motivated. At best, then, physiology motivates one basic case that can be incorporated arbitrarily into phonological rules.

Does that mean that those cases where a linguistic voicing pattern does correspond directly to outputs of the model are in fact automatic? The answer must be, only if controlling articulation in the way we have assumed is automatic. It is important that specific sets of articulatory inputs are required

to produce the outputs discussed here. The speech production system must be controlled by a real speaker in a way that ensures those inputs and no others. Possibly, as was suggested for the extrinsic vowel duration case, such control of the phonetic pattern is provided outside the phonology, with only the "exceptions" given in the phonology. But already that means that some very low-level phenomena of timing are to be included in the phonology. Consider, for example, the pattern for place of articulation to correlate with Voice Onset Time, presumably due in part to differences in the movement velocity of the various articulators. Suppose that in some language this pattern was counterexemplified by having apical stops with lower VOT values than labials, and that the reason was that the upper lip did not participate in the labial gesture (giving a lower net labial movement velocity). This would mean that in this language the place of articulation counter-pattern would be specified in the grammar, though it is concerned with mere milliseconds of timing difference. Thus, if every time we find an exception to a phonetic generalization, we state that exception in the grammar, then our notion of grammar will be much expanded. In fact, the grammar will include all the kinds of statements that remain in the phonetic component, for no kind of generalization appears to be exceptionless.

On analogy with the use of an articulatory model, we can think of preferred articulatory values as being "default" values of the articulatory system, and the outputs that result from these inputs as "default" outputs of the system. Speakers are not physically constrained to use these default inputs, and it is clear that across languages a wide variety of articulatory values are used[6]. In those cases the language has chosen to override the default settings and substitute more marked settings. Possibly the more substitutions a given output requires, the more marked it will be. Nonetheless, the default settings, where found, must still be specified at some point in the production of an utterance.

Discussion

Three candidates for inclusion in the set of phonetic universals have been considered: intrinsic vowel duration, extrinsic vowel duration, and voicing timing. None of them are automatic consequences of articulatory biomechanics, the strongest view of what a set of universals might be. None of them are necessarily universal. Thus it cannot be the case that a segmental phonetic transcription is automatically interpreted by phonetic conventions, at least with respect to such timing variables. Rather, language-specific rules extend further into phonetics than was assumed in the constrained SPE model. There are two ways that the model can be revised. If the phonetic component still consists of universals, or even just "default" cases, then almost everything is in the phonology, and the phonetic transcription will be quite narrow. If the phonetic component can include language-specific rules, then the phonetic transcription need not be so narrow, but some independent way of deciding what is in the phonetic component is needed, e.g. all timing rules. Phonetic experiments will not determine which of these possibilities is preferable. Only actually trying to devise grammars to include new phonetic data is relevant to that question.

What phonetic experiments can do is identify those parameters that must be controlled by the speaker, and default values for those parameters, by studying recurrent phonetic patterns. These patterns exist as options available to languages as physical conveniences, but not necessities. Languages must choose whether to incorporate the default, and at what level of the grammar. It is not the phonetic patterns themselves that constitute universals; rather, what are universal are the general principles that dictate the default articulatory settings.

Lindblom (1983), in discussing the concept of economy of effort as a factor in the development of sound systems, arrives at a similar overall conclusion. He stresses that speech typically underexploits the capabilities of the speech production system. In his view, more economical speech gestures are favored, but are not 'inevitable' (p. 226). Thus the occurrence and extent of consonant-vowel coarticulation, for example, may differ across speakers or be specified phonologically. Patterns found across languages are due to minimizing the expenditure of energy per unit time (p. 231); lack of a pattern in a given language indicates a greater level of performance effort of the speech system. Lindblom also concludes that the physiological mechanisms underlying economy of effort are not yet understood.

Previous approaches to language-specific exceptions to phonetic patterns have given a special grammatical role to the exceptions. Stampe (1979) proposed that a child begins acquisition with a set of phonetic processes, and replaces some of them with rules on the basis of learning. Hyman (1975) developed the idea of phonologization, that some universal phonetic processes get incorporated into the phonologies of certain languages by being made arbitrary in some way, and then playing a role in the grammar. But it seems more plausible that every aspect of phonetic control must be learned -- for example, the patterns of subglottal pressure rise and fall that give rise to consonant voicing patterns. I am suggesting here that we consider all phonetic processes, even the most low-level, to be phonologized (or grammaticized) in the sense that they are cognitively represented, under explicit control by the speaker, and once-removed from -- that is, not automatic consequences of -- the physical speaking machine.

Where this account seems unmotivated, as discussed before, are those cases where the default pattern actually occurs without exception phonetically. In these cases it would be possible to say that the default pattern is not controlled by a "phonologized" rule, but that a value is filled in by a phonetic component after all rules have applied. Consider, however, such a phonologization account of extrinsic vowel duration in various languages. That account will distinguish languages like Russian and German (with final devoicing and opaque vowel length differences) from languages like French (with phonetically transparent vowel length differences). Russian and German will have phonologized vowel length, while French will not; it will have durations supplied by the phonetics. Suppose now that French acquires a rule of final consonant devoicing like that of German or Russian, and that, as in German and Russian, vowel length is sensitive to phonological voicing. The phonologization account would have to say that at the moment the devoicing rule is added to the French grammar, vowel length also becomes a grammatical rule, as opposed to a default option or pattern. Since the only change in the vowel length pattern is that it has changed from phonetically transparent to opaque, then rule transparency must be criterial in assigning phonetic patterns to the phonetics or the phonology. On the other hand, if all phonetic patterns, including transparent vowel length, are represented in the grammar, then the only change in the French grammar is the addition of the devoicing rule. In the absence of arguments for the transparency criterion, then, the phonologization -- unmarked patterns not in the grammar -- account seems more complex than required.

The view that all phonetic phenomena are controlled by rule has a further interesting implication. Anderson (1981) argued that phonological rules by definition aren't natural -- they're what's left over when everything else is factored out. As a response to various theories of 'natural phonology', this argument is valid. But it leaves the frequent phonetic naturalness of rules --

even rules with exceptions on the surface -- unexplained. It sounds ad-hoc that some rules (most low level ones) should actually be natural, while other rules (the opaque ones) only look natural. But once we recognize that all phonetic patterns are rule-governed, and once-removed from the physical machine, then naturalness can be seen as a more abstract and general property of rules, wherever they are in the phonology. Various rules will have in common the fact that they embody default patterns. Some of these rules will apply transparently; others will apply opaquely. Naturalness is not directly a fact about the speaking machine. It is a fact about the phonological component: the phonology values highly rules that in form indulge the preferences of the speaking machine.

Patterns of phonetic detail are interesting, then, not because they constitute a special universal component outside of grammars, one whose workings are quite different from those of phonology, but rather because they are an intergral part of phonology. It seems likely that there are no true linguistic phonetic universals, and that a language's grammar controls all aspects of phonetic form.

Acknowledgements

Preparation of this report was supported by a grant from NIH to Peter Ladefoged. Earlier versions of this work were presented as talks to helpful groups at the University of British Columbia, the University of California at Irvine, the 1983 International Congress of Phonetic Sciences, and the Phonetics Lab at UCLA. I would especially like to thank Peter Ladefoged, Vicki Fromkin, Marie Huffman, Bruce Hayes, and John Ohala for comments on the manuscript, and John Westbury for our collaborations.

Footnotes

1. Also taken into account at this stage are nongrammatical suprasegmental parameters, both for languages (base of articulation) and for individuals at a given moment (voice quality, rate of utterance).

2. It will not matter for this discussion whether the pattern is seen as shortening of vowels before voiceless obstruents, or lengthening of vowels in converse environments.

3. Chen's comparisons confounded language and position of the vowel+consonant in a word: some languages were represented mainly by monosyllables, others mainly be disyllables with a medial vowel+consonant. The degree of vowel duration difference is known to vary even within a single language according to position (compare Sharf 1962 and Klatt 1973 with Lehiste 1970).

4. Though the rule of devoicing does not guarantee that the neutralized consonants themselves are identical; cf. Dinnsen (1982).

5. Ohala links final devoicing to an observed lengthening of final consonants, that is, they devoice for the same reason geminates do. Notice that in our modeling we have not lengthened final consonants, showing that such lengthening is not required, though of course it would have the enhancing effect Ohala describes.

6. Though these settings have some absolute limits in the physical world (e.g. how fact the tongue can move), it is interesting that these limits are not typically approached in speaking. For example, the changing volume of the oral cavity is relevant in any consideration of voicing maintenance for stop consonants, as we have seen, and obviously there is some finite limit on how large an individual's oral cavity can become. But this limit is probably never approached in speaking; Westbury (1983) shows that the set of possible maneuvers to expand the oral cavity makes so much expansion possible that from the point of view of speaking the oral cavity seems to have unlimited potential volume. When a speaker exploits these maneuvers is a separate question, of course.

References

Anderson, S. (1974). The Organization of Phonology. New York: Academic Press.

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

Chen, Matthew (1970). "Vowel length variation as a function of the voicing of consonant environment." Phonetica 22. 129-59.

Chomsky, N. (1964). "Current issues in linguistic theory". In The Structure of Language, J. A. Fodor and J. J. Katz, eds., pp. 50-118. Englewood-Cliffs, NJ: Prentice Hall.

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

Dinnsen, Daniel A. (1980). "Phonological rules and phonetic explanation". J. Linguistics 16. 171 - 191.

Dinnsen, Daniel A. (1982). "Abstract phonetic implementation rules and word-final devoicing in Catalan". Talk presented at the annual meeting of the Linguistic Society of America, San Diego.

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

Fowler, C. A. (1980). "Coarticulation and theories of extrinsic timing". J. Phon. 8.113-133.

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

Fromkin, V. A. (1977). "Some questions regarding universal phonetics and phonetic representations". In Linguistic Studies offered to Joseph Greenberg on the occasion of his sixtieth birthday, ed. A. Jilland, pp. 365-380. Saratoga, Calif.: Anma Libri and Co.

Fujimura, O., and J. Lovins (1978). "Syllables as concatenative phonetic units". In Syllables and Segments, ed. A. Bell and J. Hooper, pp. 107-120. Amsterdam: North-Holland Publishing Co.

Hooper, J. B. (1976). An Introduction to Natural Generative Phonology. New York: Academic Press.

Houlihan, K. (1982). "Is intervocalic voicing a natural rule?" Talk presented at the annual meeting of the Linguistic Society of America, San Diego.

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

Keating, Paticia A. (1979). "A phonetic study of a voicing contrast in Polish." Unpublished Ph.D. dissertation, Brown University.

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

Klatt, D. (1973). "Interaction between two factors that influence vowel duration". J. Acoust. Soc. Am. 54. 1102-04.

Ladefoged, Peter (1971). Preliminaries to linguistic phonetics. Chicago: University of Chicago Press.

Ladefoged, P. (1980) "What are linguistic sounds made of?" Language 56(3). 485-502.

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

Lindblom, B. (1967). "Vowel duration and a model of lip mandible coordination". STL-QPSR 4. 1-29.

Lindblom, B. (1983). "Economy of Speech Gestures". In The Production of Speech, ed. P. F. MacNeilage, pp. 217-246. New York: Springer-Verlag.

Lisker, L. (1957). "Closure duration and the intervocalic voiced-voiceless distinction in English". Language 33. 42-49.

Lisker, Leigh, and Arthur S. Abramson (1964). "A cross-language study of voicing in initial stops: Acoustical measuremnts." Word 20.384-862.

MacNeilage, P. (1970). "Motor control of serial ordering of speech". Psych. Rev. 77. 182-196.

Maddieson, Ian (1983). Patterns of Sounds. Forthcoming from Cambridge University Press.

Müller, E. M., and W. S. Brown, Jr. (1980). "Variations in the Supraglottal Air Pressure Waveform and their Articulatory Interpretation". In Speech and Language: Advances in Basic Research and Practice, Vol. 4, ed. N. Lass, pp. 317-389. New York: Academic Press.

Nelson, W. L. (1980). "Performance bounds in speech motor control of jaw movements". J. Acoust. Soc. Am., 68, Suppl. 1. S32 (A).

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

Pierrehumbert, Janet Breckenridge (1980). "The Phonology and Phonetics of English Intonation." Unpublished Ph.D. dissertation, M.I.T.

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

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

Postal, Paul (1968). Aspects of phonological theory. New York: Harper and Row.

Rothenberg, M. (1968). The Breath-Stream Dynamics of Simple-Released-Plosive Production. Bibliotheca Phonetica 6. Basel: S. Karger.

Sharf, D. (1962). "Duration of post-stress inter-vocalic stops and preceding vowels". Lg. Speech 5. 26-30.

Stampe, D. (1979). A Dissertation on Natural Phonology. 1973 Ph.D. dissertation distributed by IULC.

Westbury, J. R. (1979). Aspects of the Temporal Control of Voicing in Consonant Clusters in English. Ph.D. dissertation, U. Texas, published as Texas Linguistic Forum 14. 1-304.

Westbury, J. R. (1983). "Enlargement of the supraglottal cavity and its relation to stop consonant voicing". J. Acoust. Soc. Am. 74(4). 1322-36.

Westbury, J. and Keating, P. (1980). "Central representation of vowel duration". J. Acoust. Soc. Am. 67, Suppl. 1, S37 (A).

Westbury, J. and Keating, P. (1984). "On the naturalness of stop consonant voicing". Manuscript in preparation.

Vowel allophones and the vowel-formant phonetic space

Patricia Keating, Marie Huffman, and Ellen Jackson

Paper presented at ASA meeting, November 1983, San Diego

It is well-known that vowels vary phonetically across segmental and prosodic contexts, as shown years ago, for example, by Stevens and House (1963), Ohman (1966), and Lindblom (1963), among others. In English, quite a range of segment types result from such context effects: nasalized vowels, long vowels, central vowels, front rounded vowels. English vowel tokens thus exploit the phonetic space much more than a list of the vowel phonemes would suggest. Our research project is concerned with the extent to which languages share such allophonic variation, with focus on the F1-F2-F3 phonetic space. If allophones tend to fill this space, then the interesting possibility arises that languages with quite different vowel phonemes could nonetheless be rather similar phonetically. On the other hand, it seems more likely that a language's phonemic inventory and structure determine its phonetic variants. Not only which vowels it has, but also what consonant and prosodic contexts those vowels occur in should matter. Yet another factor affecting variation could be differences among vowels in how they vary. Recently Louis Goldstein suggested, based on articulatory modeling and language change, that vowels should differ in where they spread in the vowel formant space: front vowels spread up while back vowels spread forward as well as up. This proposal has not been tested against patterns of natural allophonic variation.

In this paper we will discuss allophonic variation in the F1-F2 space for Japanese and, briefly, Russian, vowels. Both of these languages have five vowels. Japanese has short or long monophthongal /i e a o ɯ /. /ɯ / is a high back vowel that differs from an [u] in that it is not rounded and is perhaps more centralized, both of which result in a higher second formant. Another difference between Japanese and English or Russian is that Japanese is a pitch-accent, not a stress, language, meaning that each vowel has a high or low tone according to fixed word patterns. Japanese, then, does not have vowel reduction in unstressed syllables as English and Russian do. Instead, Japanese short low-tone vowels in certain environments are shortened to the point of being inaudible, or are deleted altogether. This rather different sort of prosodic system makes it plausible that Japanese vowels should vary less across contexts than vowels in other languages. However, in our data we have found that Japanese vowels do vary, and nearly fill the F1-F2 space, although the [u] region remains unfilled.

Our data consist of LPC-measured formant frequencies for vowel tokens of seven young male speakers of Tokyo Japanese. Table 1 shows the three word lists that each speaker read twice. In addition, each speaker read several prose texts. For each vowel token, a single set of F1, F2, F3 measurements was obtained by averaging LPC values for a 30 to 40 msec span in the middle of the steadiest portion.

For the prose, we attempted to get 100 tokens of each vowel from each speaker, but some speakers deleted /i/ and /ɯ/ so freely that we could only get about 80 tokens of each of these vowels. For each speaker, equal numbers of the 5 vowels were used. Vowel sequences and nasalized vowels were not included.

i	hibi	tabi	wasabi
e	hebi	nabe	otabe
a	habu	kaba	anaba
o	hobo	yabo	otsubo
ɯ	hubo	kobu	manabu
tones	H L	H L	L H H

Table 1

When the 30 word-list tokens for each speaker are compared, the results are similar, even in absolute frequency. Figure 1 shows roughly where the 5 vowels are located for the entire group. The axes have been flipped so that the vowels are arranged as in a traditional tongue-position vowel chart. Note that /e/ and /o/ are about equal in height, while /ɯ/ and /a/ are about equal in backness. /i/ is tacked onto this diamond pattern, being higher and fronter. The space has a gap where most languages have an [u] or [ʊ]. The next two figures show the set of values for two speakers each representative of a subset of the group. Figure 2 shows a speaker whose mid and high vowels are quite separated. Figure 3 shows a speaker whose mid and high vowels are less separated.

Now let's see what happens to the vowel space with the tokens from the uncontrolled prose text, for these same two speakers. Figure 4 is the first word list set we saw in Figure 2, overlaid with the prose tokens. This speaker, the one with the more separated word-list vowels, has relatively less variation for each vowel in the prose condition. The ellipses are small, they overlap little, and the center of the space is empty. Figure 5 shows the word list and prose tokens of the speaker whose word list vowels are closer together. His vowel ellipses are larger and they overlap more, leaving no gaps. Furthermore, this space is larger than any other speaker's.

For the 7 speakers, /ɯ/ varies the most: it covers a larger area than other vowels for each speaker's space, and its position in the space differs more across speakers. Figure 6 shows just the /i/ and /ɯ/ tokens for the speaker whose vowels spread the least. As you can see, the /ɯ/ varies more than the /i/, with some overlap between them. Even so, the /i/ is still clearly higher than the /ɯ/. Though the /ɯ/ spreads back into the [ʊ] region, it does not spread up into the high back rounded [u] corner. In some cases, tokens of /o/ do spread up into that corner, but not often. Figure 7 shows a representative /ɯ/ and /o/. /o/ is clearly more back than /ɯ/, and sometimes it is as high--that is, like an [ʊ] vowel. But overall the [u] area remains empty.

Is there any evidence in these data that vowels differ in their directions of spreading, as Goldstein suggested? We find no consistent differences for the seven speakers, except perhaps that /i/ varies the least and /ɯ/ varies the most, but never spreads up and back in the space. In addition, we find no effects of tone on vowels' spreading in these data. High and low tone vowels spread alike.

Let us now compare the situation in Japanese with that in Russian. Russian also has five vowels, with /u/ instead of /ɯ/. It also has stress and vowel reduction, and a contrast of palatalized vs. plain consonants that is said to produce much variation in vowels. For example, the Russian /i/ has an allophone [ɨ] not too unlike Japanese /ɯ/--but Russian also has /u/. How do the 5 Russian

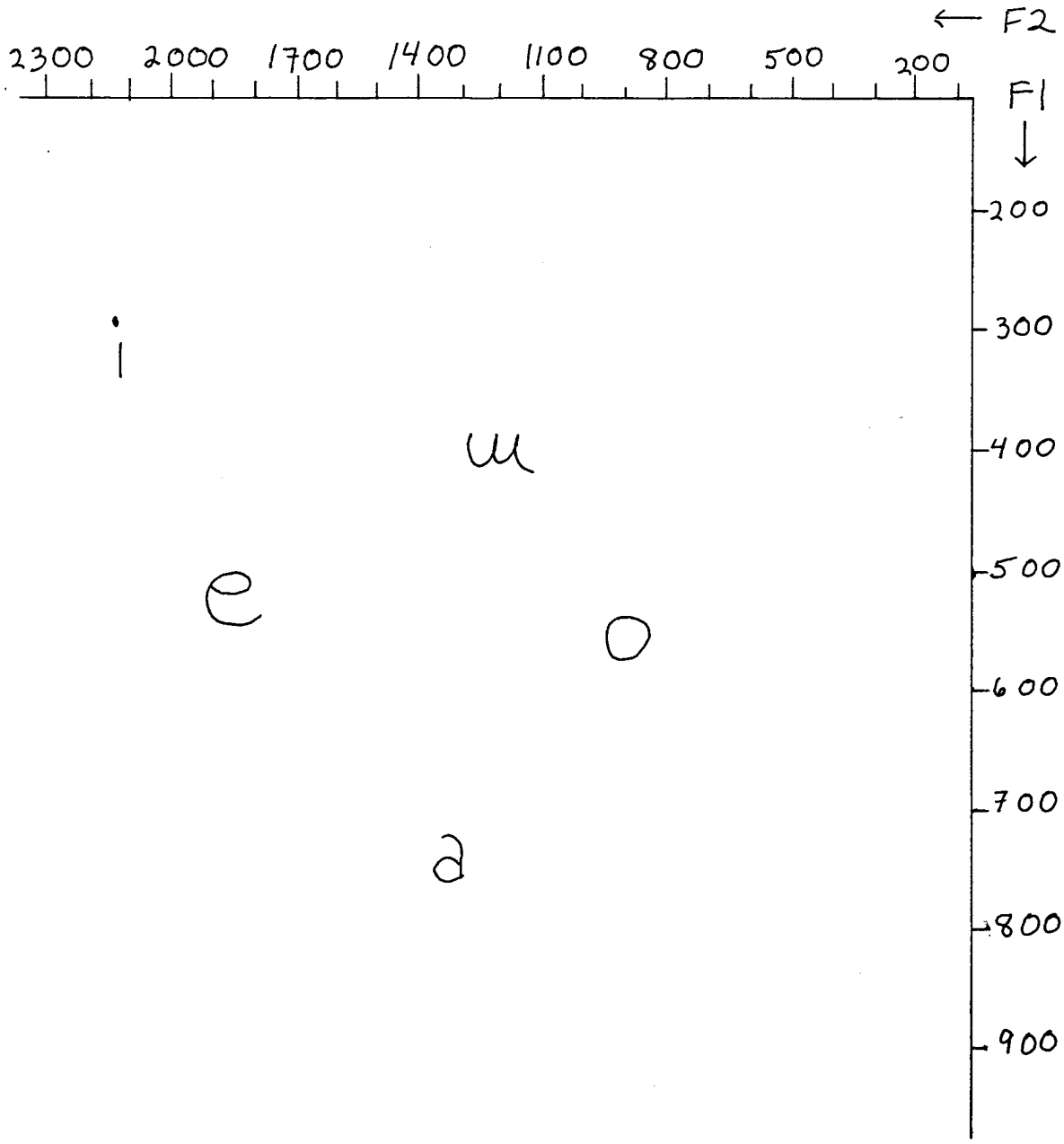

← F2
2300
2000
1700
1400
1100
800
500
200
F1
200
300
400
500
600
700
800
900
i
u
e
o
a

Figure 1

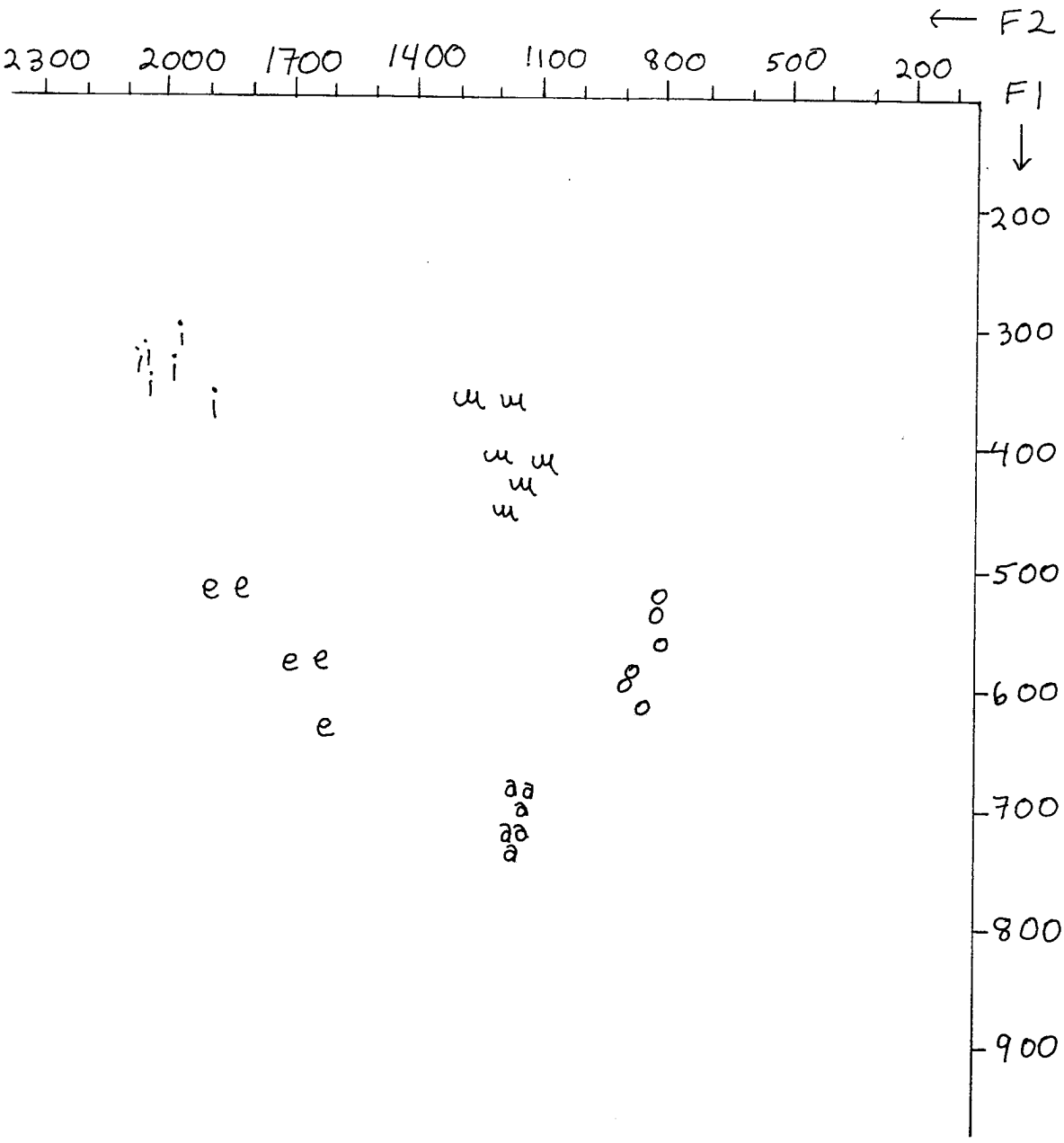

← F2
2300
2000
1700
1400
1100
800
500
200
F1
200
300
400
500
600
700
800
900

Figure 2

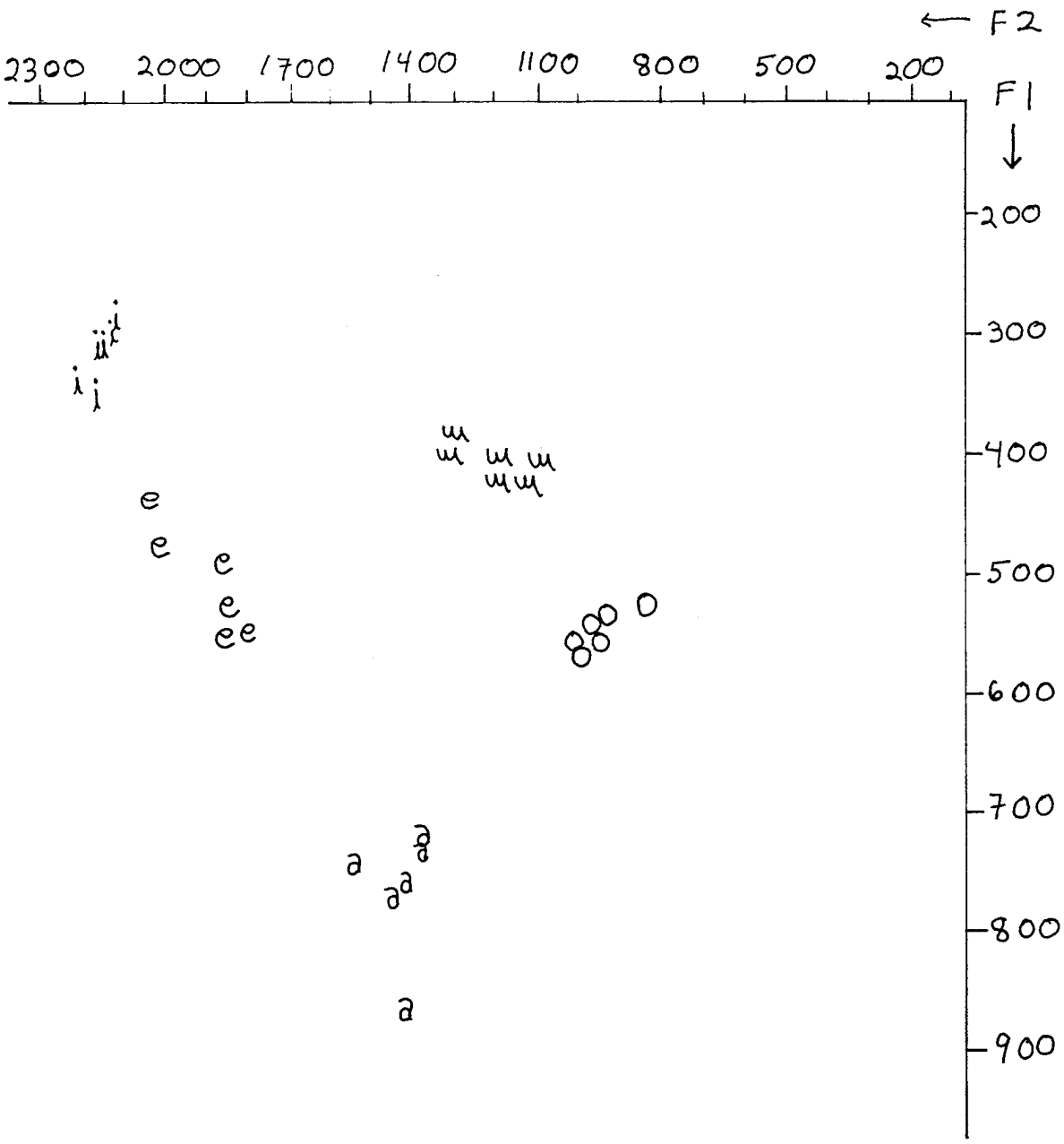

← F2
2300
2000
1700
1400
1100
800
500
200
F1
200
300
400
500
600
700
800
900

Figure 3

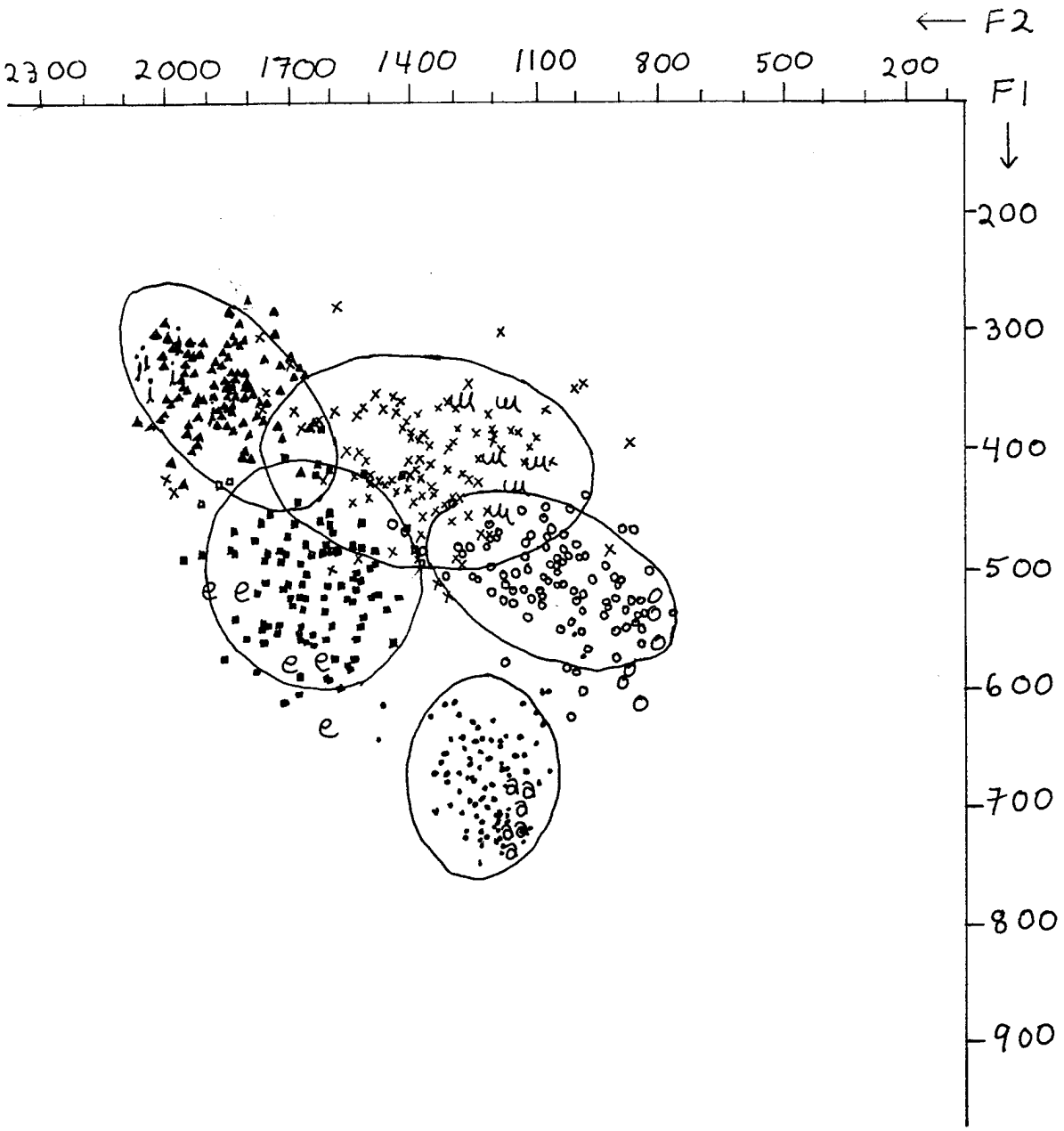


Figure 5

Word list = Printed letters

Prose symbols:

i = ▲
e = ■
a = ·
o = o
ɯ = ×

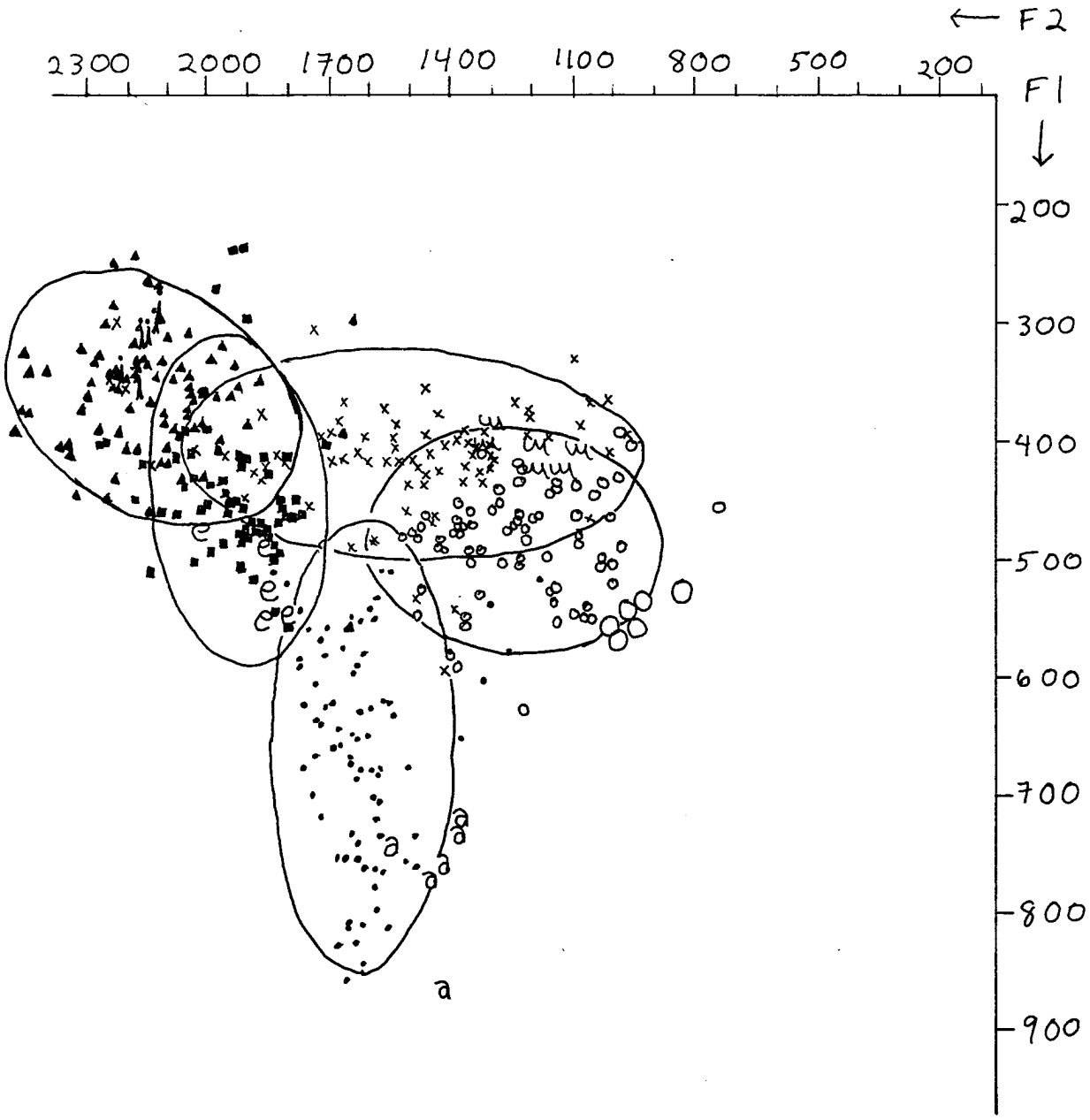


Figure 5

Word list = Printed letters

Prose symbols:

i = ▲
e = ■
a = ·
o = o
ɯ = x

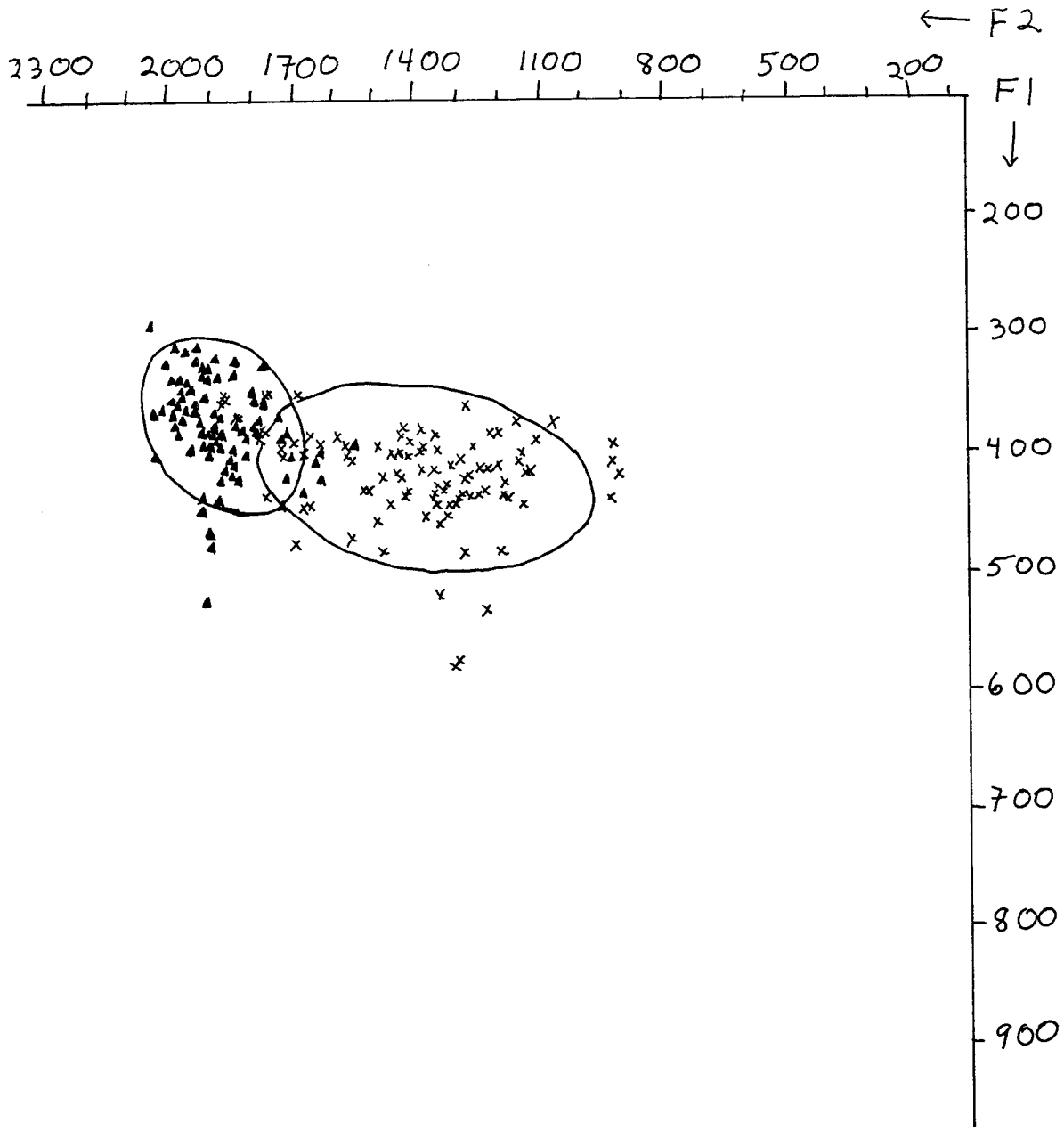


Figure 6

Prose symbols:

i = ▲

ɯ = x

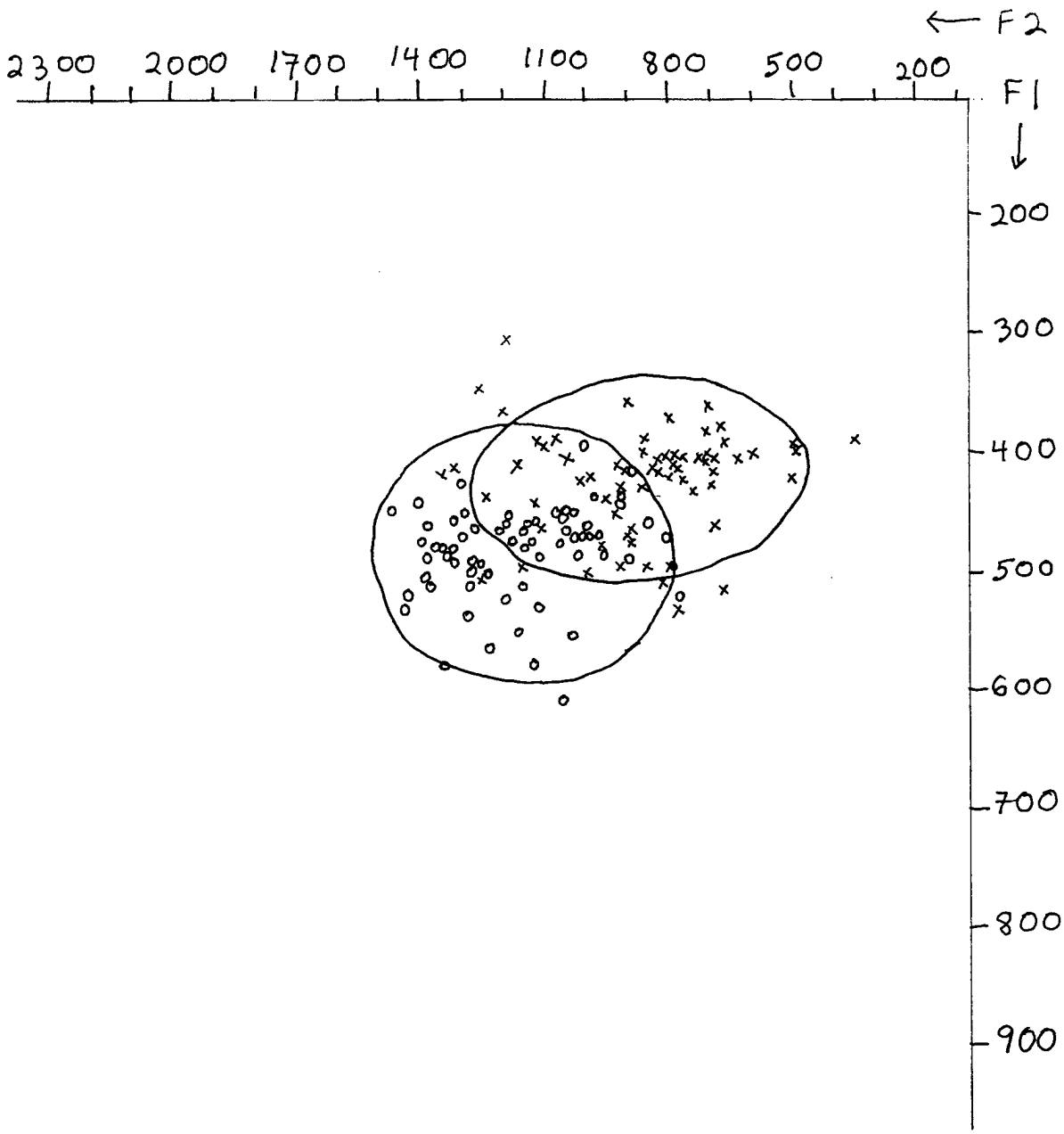


Figure 7

Prose symbols:
o = o
ɯ = x

vowels compare with Japanese? So far we only have partial data for 1 speaker from Kiev. Figure 8 shows measurements of his word-list vowels in labial consonant contexts similar to the Japanese words, except for the palatalization contrast. The dotted circle indicates an area of vowel tokens that had to be measured from spectrograms--the LPC routine couldn't spearate F1 from F2. Note that in Russian the /i/ and /u/ are equally high, and that the variation in /i/ represents the 2 allophones due to consonant differences. Now consider the Russian prose tokens in Figure 9. Compared with Japanese, the Russian speaker shows more variation for each vowel, and more overlap of vowels. Compared with his word-list tokens, the Russian speaker's prose tokens spread inward. Is this because unstressed tokens in the text are reduced and centralized? No, the stressed prose tokens considered separately also fill the space. The unstressed tokens do not cover new parts of the space; they simply concentrate in the most centralized areas of the stressed vowels. The large amount of variation and overlap, then, must arise from the consonantal effects. For example, palatalized coronal consonants have the effect of fronting and raising /i/ tokens while eliminating back /u/ tokens. The result is a vowel space quite similar to that of Japanese.

In conclusion, we have found that even languages with few vowels will essentially fill the vowel space with vowel tokens in running speech. Context effects will not, however, produce an [u] in Japanese, leaving a gap in the space reflecting the phonemic inventory. Further, the spreading of vowels across contexts apparently does not depend on prosodic effects, either tone or stress. Languages <u>are</u> perhaps more similar phonetically than we might have expected, but differences reflecting the phonology are also apparent.

Acknowledgement

This research was supported by a grant from NSF to Peter Ladefoged.

References

Goldstein, Louis (1983). Vowel shifts and articulatory-acoustic relations. Abstracts of the Tenth International Congress of Phonetic Sciences. A. Cohen and M. P. R. v.d. Broecke, eds. (Foris).

Lindblom, Björn (1963). Spectrographic study of vowel reduction. J. Acoust. Soc. Am. 35: 1773-1781.

Öhman, S. (1966). Coarticulation in VCV utterances: spectrographic measurements. J. Acoust. Soc. Am. 39: 151-168.

Stevens, K. N., and A. S. House (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. JSHR 6: 111-128.

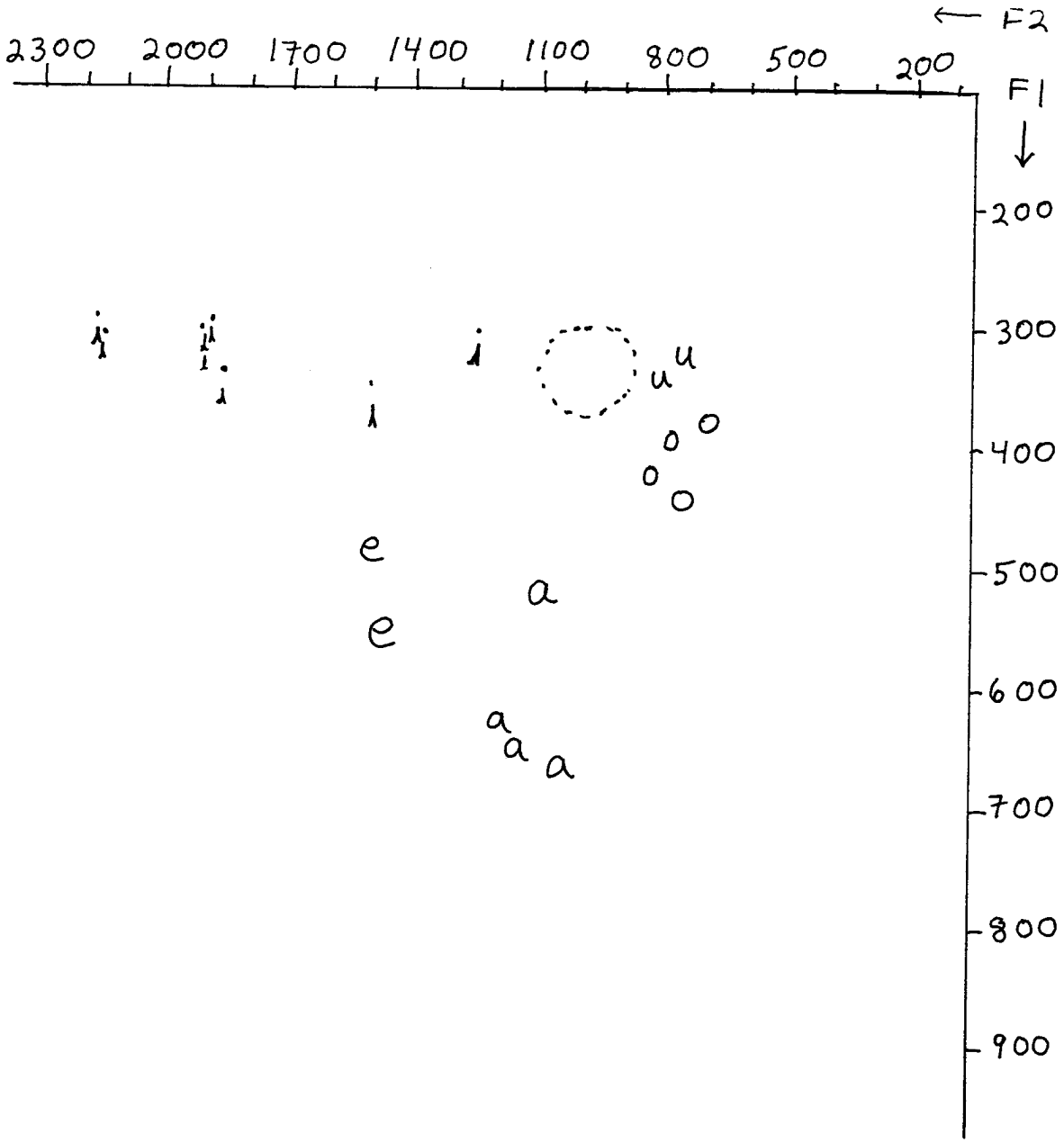

← F2
2300
2000
1700
1400
1100
800
500
200
F1
200
300
400
500
600
700
800
900

Figure 8

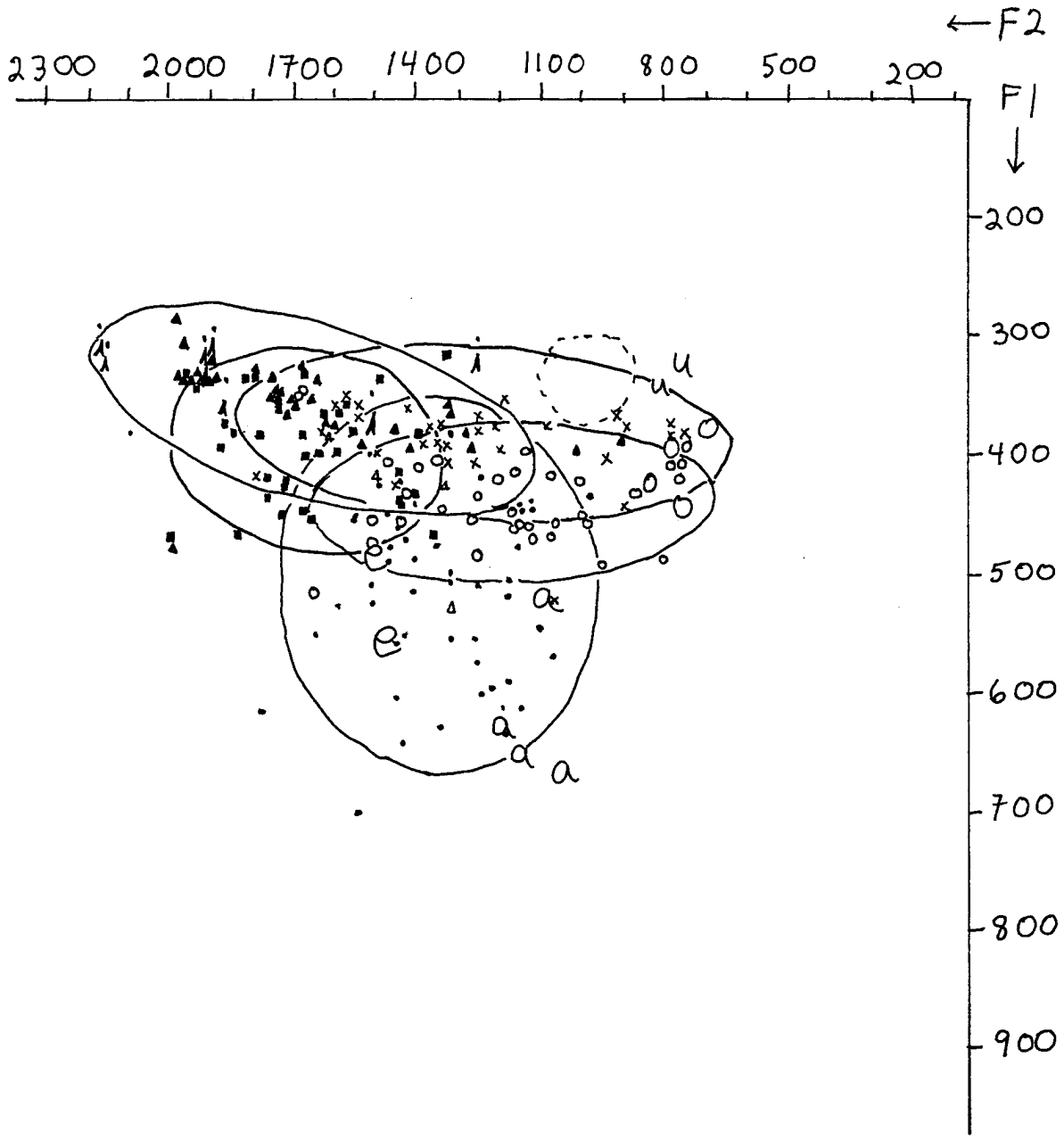


Figure 9

Word list = Printed letters

Prose symbols:

i = ◢
e = ■
a = ·
o = o
u = x

Places of Articulation:
An investigation of Pekingese fricatives and affricates

Peter Ladefoged

Phonetics Laboratory,Department of Linguistics, UCLA,Los Angeles, CA 90024, USA.

and

Zongji Wu

Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China.

Phoneticians generally believe that there is a set of places of articulation. This notion is embodied in charts of consonants, such as that of the IPA, where the separate columns specify separate places of articulation. It is also implied in every system of distinctive features (e.g. Chomsky and Halle, 1968). Of course, most phoneticians agree that the boundaries between adjacent categories are not clearly determined. They accept, for example, that there is no precise boundary between palatal and palatoalveolar sounds, so that it is hard to say whether the Akan word for "father" should be transcribed as [aɟa] or [ad̠a], with a palatal or a palato-alveolar stop. Or, to take another example, they know that the distinction between dental and alveolar stops is somewhat tenuous, and that the IPA recognizes this fact by placing these sounds in a single, undifferentiated column. (The reason in this latter case is presumably simply a historical accident. The IPA was founded by Europeans speaking languages that do not distinguish between dental and alveolar stops. If it had been founded by Australian aboriginals we might have had a different chart.) But, after allowing for the fact that the boundaries between adjacent categories cannot always be clearly delineated, it is usually held that consonants can be described in terms of a set of specific places (and manners) of articulation. They are not treated like vowels, which are clearly recognized as being points in a space. Consonants are regarded as belonging in discrete cells on a chart.

Some of the major problems in this notion occur in the description of fricatives. The IPA chart has a larger number of symbols in this category than in any other. Phoneticians have long recognized that there are subtle differences between the fricatives of different languages that necessitate the use of different symbols. But the values of these different symbols have not always been clear. In order to shed some light on part of this problem, we will present some data on fricatives and affricates in Standard Colloquial Chinese (Pekingese) as spoken in Peking. We will refer to this language as Pekingese.

The set of sounds we will consider are given in an IPA transcription in Table 1, the traditional IPA category labels are also shown.

Table 1. The fricatives and affricates of Pekingese

	Labiodental	Alveolar	Retroflex	Alveolopalatal	Velar
voiceless fricative	f	s	ʂ	ɕ	x
voiceless affricate		ts	tʂ	tɕ	
aspirated affricate		ts^h	$tʂ^h$	$tɕ^h$	

Words illustrating the sounds in Table 1 are given in the official Pinyin spelling in Table 2. It may be seen that this romanization differs in many ways from IPA practices.

Table 2. Words illustrating the sounds in Table 1, written in the official Pinyin romanization

fa	sa	sha	xia	ha
	za	zha	jia	
	ca	cha	qia	

Procedure

The data to be reported here have been selected from those in a much larger study (Wu, 1963), which is designed to illustrate the principal features of all the sounds of Pekingese. The complete set of data includes palatograms, audio recordings and various kinds of acoustic analyses for five subjects, together with x-ray photographs of three of the subjects. We will be concerned here mainly with the articulatory data for the three principal subjects, with only occasional comments on the acoustic data. All of the subjects were native speakers of Pekingese, born and raised in the neighborhood of Beijing. Speaker A was a male. Speakers B and C were females. All of them were around 20 years old.

When producing the palatographic data, subjects said each of the words in Table 2 in citation form. The palatograms were obtained using a modified version of a technique described by Hammarstrom (1957). Chinese carbon black ink was painted onto the tongue. The subject then said the word being investigated, causing the ink from the tongue to be deposited on the upper surface of the mouth. A record of the contact area was made by placing a mirror in the mouth so that this surface could be photographed. The enlargements from the negatives were all made exactly life size, as checked by comparison with an artificial palate made from a dental impression of each subject's mouth. Tracings of the contact areas were made of the outlines for each utterance.

Lateral x-ray photographs were taken. In order to enhance the outline of the vocal tract, the nose, the lips, and the center line of the tongue were painted with a barium solution. Further enhancement was achieved when making the life-size photographic prints from the negatives. Parts of the image were masked by an adjustable cut-out with a shape corresponding to the bony structures of the upper and lower jaw. This mask was dodged (held, but with a slight movement) between the lens of the enlarger and the photographic papers, so that the images of the lips and the oral cavity were fully exposed without the bony structures becoming over-exposed.

In addition, instead of the usual way of tracing the outline of the vocal tract by placing tracing material on top of the photograph (a practice that inevitably lessens the visibility of the image), a new technique was evolved. The positions of the articulators and relevant structures such as the surfaces of the teeth were drawn directly on the photograph in black waterproof ink. The image on the photograph was later leached away with a potassium permanganate solution, leaving only the required drawing.

The x-ray photographs were taken while the subject was trying to maintain the articulatory posture for the initial consonant in a syllable (the closure phase in the case of the affricates). Consequently the data may not provide a valid description of the sounds as they occur in natural speech. However the general consistency both across the three subjects, and within each subject for the pairs of sounds that differ only in aspiration, together with the fact that the x-rays show similar points of contact to those on the palatographic data (which were obtained from pronunciations of whole words that sounded completely natural) all tend to confirm the validity of the x-ray data.

Results

Data based on the X-rays of the fricatives [s,ʂ,ɕ] are shown in Figure 1. Corresponding palatographic data are shown in Figure 2. The first point to note is that for all three sounds for all three subjects the upper and lower teeth are fairly close together. It is this narrow channel between the teeth that gives all these sounds their sibilant (strident) quality. In each of the sounds the tongue forms a differently shaped channel for the air. But the main source of acoustic energy is always the turbulence that arises when this air passes between the nearly clenched teeth. This similarity in the articulation of all three sounds is not captured in any way by the traditional IPA categories. But it is, of course, explicitly recognized by distinctive feature systems such as that of Jakobson, Fant and Halle (1951) by the provision of the feature Strident.

As the top row of Figure 1 shows, all three subjects produced [s] with the tip of the tongue; and in all three cases there is a hollowing of the tongue such that the tongue is concave with respect to the roof of the mouth. Again we have a failure of the descriptive categories, this time in the case of both the IPA and the feature system approaches. If we categorize this sound simply as Alveolar (or [+ coronal, - anterior]) we are implying that the tongue shape is the same as in other alveolars such as [t]. But [t] does not have this hollowing of the tongue. It would seem as if we have to make our phonetic definitions context sensitive: if Alveolar and Fricative (or [+ coronal - anterior] and [+ strident]), then a hollowing of the tongue is implied.

It is interesting to try to estimate the size of the channel when the constriction is at its greatest. The palatograms show that subjects B and C make this sound with a narrow slit, with width of 4.5 mm for subject B and 3.75 mm for subject C. (Subject A makes this sound with the narrowest channel on the teeth, so palatographic data is not available for this measurement.) The height of the slit is about 1 mm for subjects A and B, and even less for subject C.

The position of the point of greatest constriction is sightly different for each subject. For subject A it is on the teeth for Subject B slightly behind the teeth, and for subject C still further back on the alveolar ridge. Given these data, it seems that the exact place of articulation (in the sense of the precise location of this channel) is not particularly critical for these sounds. What matters more, as in many sounds, is the shape of the vocal tract as a whole.

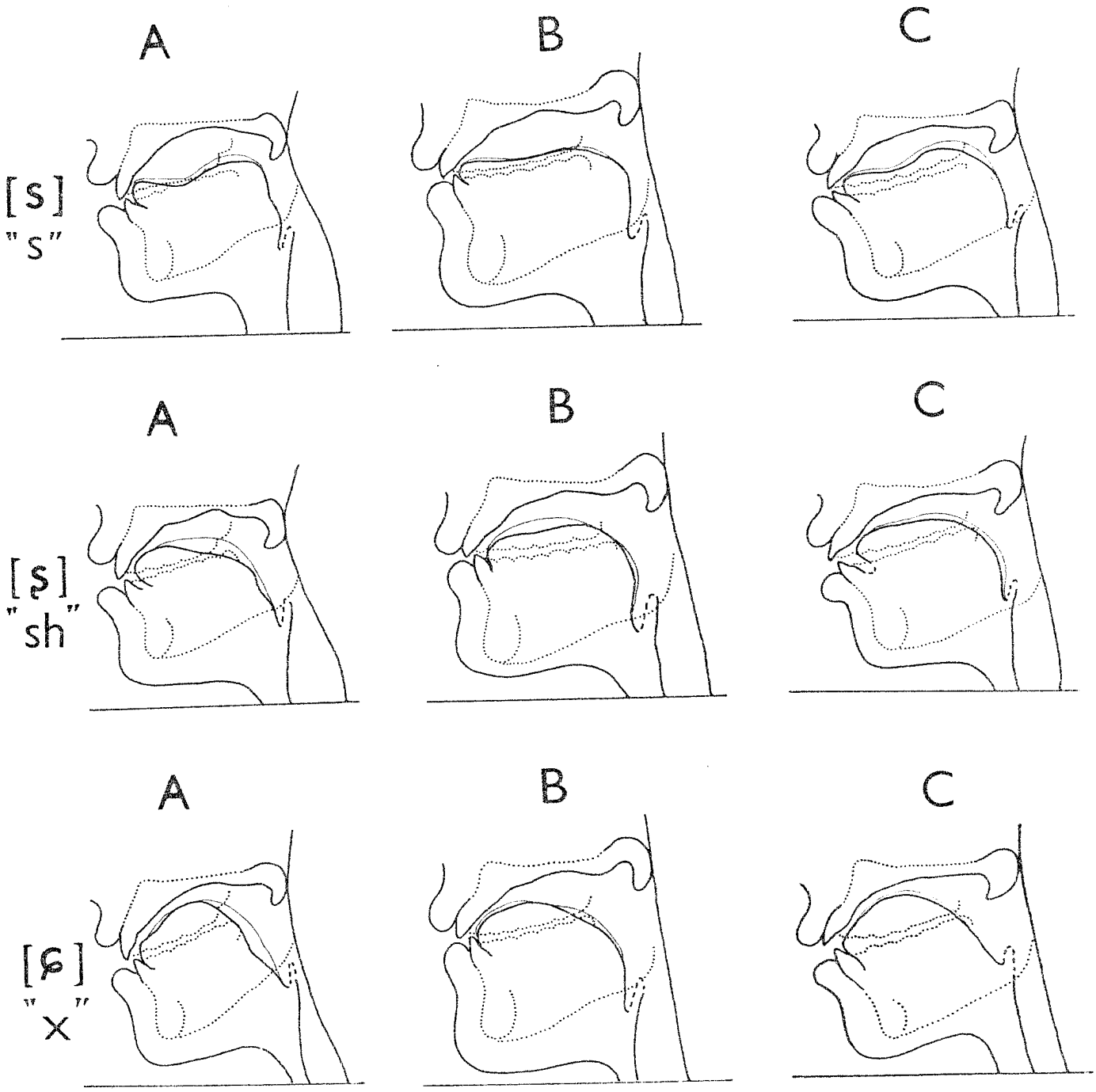


Figure 1. Tracings from x-rays of three speakers producing Pekingese sibilant fricatives. Where there are two lines drawn for the tongue, the lighter line represents the positions of the sides of the tongue.

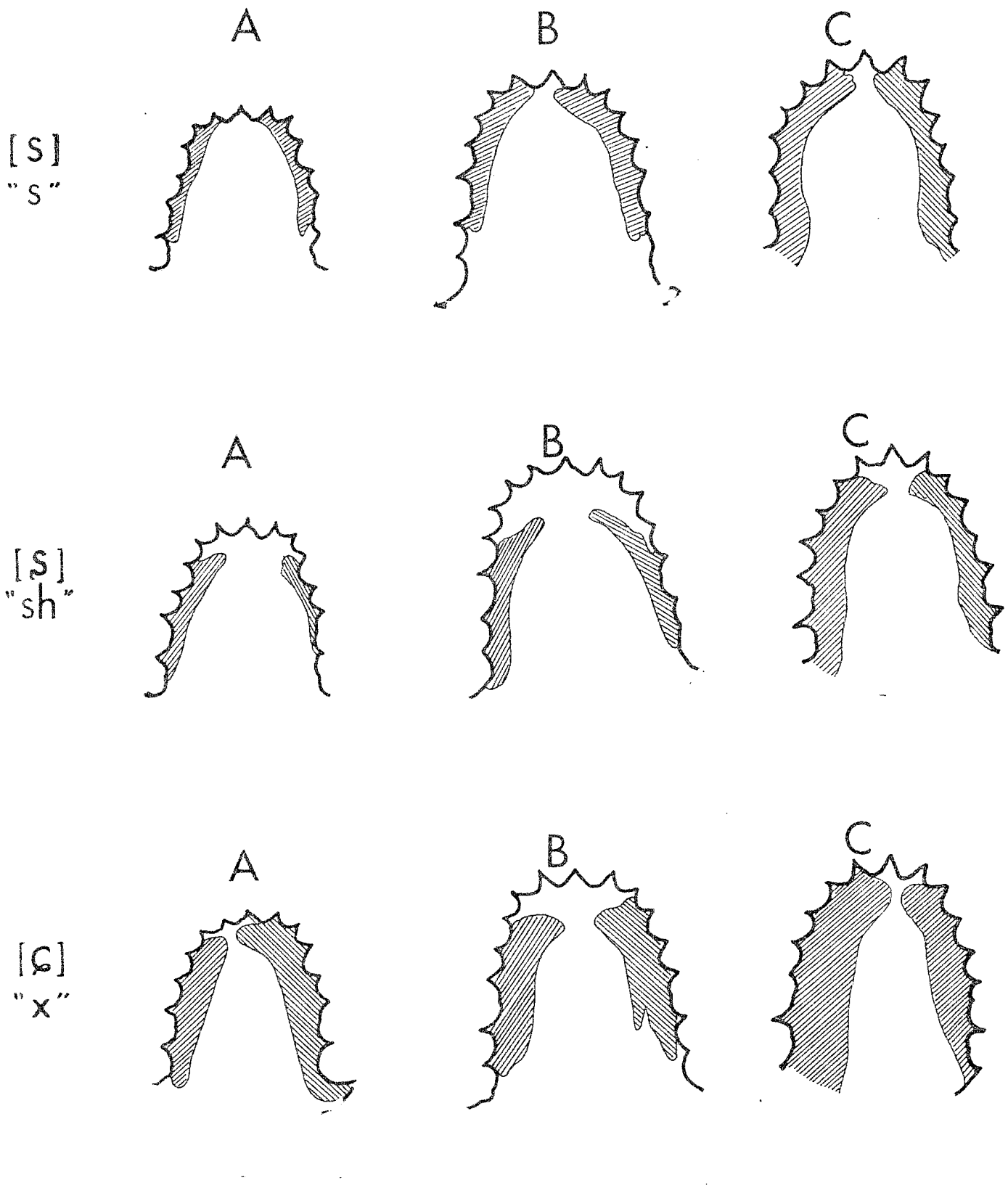


Figure 2. Palatograms of three speakers producing Pekingese sibilant fricatives.

The so-called retroflex [ʂ] is shown in the middle rows of Figures 1 and 2. It is immediately apparent that this sound does not have the tip of the tongue curled up and backwards, as it does in Indian sounds symbolized in a similar way (compare Balasubramanian, 1972). All three speakers produce the constriction for this sound with the upper surface of the tip of the tongue (as compared with the under surface of the tip, as is common in the Dravidian languages of Southern India). The constriction is at about the same place for all three speakers, namely at about the center of the alveolar ridge. Both the height and the width of the channel are greater than in [s]; but the width varies considerably, from 18.5 mm for subject A to 5 mm for Subject C. The increased channel size results in the jet of air not attaining as high a velocity as it does in [s].

The articulatory differences between [s] and [ʂ] in the width of the constriction are not captured by any of the traditional descriptions in terms of place of articulation. Nor are the differences in the shape of the rest of the tongue. For [ʂ] the front of the tongue is fairly flat for subjects A and C, and only slightly hollowed for subject B. The root of the tongue is more advanced in [ʂ] than it is in [s] for all three subjects, resulting in the tongue being more bunched up lengthwise. These aspects of the articulation are not described simply by specifying the place of articulation as retroflex, or indeed, by any description that merely specifies the location of the source of the fricative turbulence.

The tongue has a very different position in [c] from that in either of the other two sounds we have been considering, as may be seen from the data in the bottom rows of Figures 1 and 2. It is much higher in the mouth, forming for both subjects, a comparatively long, flat, constriction. Subject A probably produces this raising of the front of the tongue by the action of the genioglossus muscle. As can be seen in Figure 1, the sides of the tongue (the thinner line) are higher than the center of the tongue (the solid line) in the region just above the epiglottis. This is caused by the genioglossus muscle pulling the root of the tongue forwards and producing the characteristic groove in this region that results from the action of this muscle. Subject B and C do not have such a deep hollowing of the root of the tongue, and may produce the raising of the body of the tongue by the action of the mylohyoid muscle. Use of these different muscles would account for the differences in the vocal tract shapes of the subjects.

The narrowest channel occurs near the front part of the alveolar ridge for subject C, and notably further back for subject B. For neither subject is the constriction in the same place as in either of the other two sounds; it is farther back than in [s], but not quite as far back as in [s]. Accordingly, although the high position of the front of the tongue might lead to this sound being considered as palatalized, it is in no sense a palatalized version of either of the other two sounds. It must be assigned to a separate "place of articulation" as has been done by the IPA. And, as in the previous cases, the category to which this sound is assigned has to specify not only the location of the constriction but also the shape of the whole body of the tongue. (Or, alternatively, we have to add some new low level phonetic features to take care of these differences in tongue shapes.)

This category may have to be interpreted in other ways when used to describe sounds in other languages (just as we have already noted that the category retroflex has to be assigned a different interpretation when used to describe Dravidian languages). The IPA (1949) regards the Pekingese sound as the same as

the Polish "ś" as in "geś". But both our own listening and the x-ray data in Puppel (1977) indicate that the Polish sound is more palatal.

The affricates of Pekingese are illustrated in Figures 3-5. Figures 3 and 4 show show that the unaspirated affricates [ts, tʂ, tɕ] and the aspirated counterparts [tsʰ,tʂʰ,tɕʰ]; are fairly similar to each other; there do not seem to be any systematic differences. Furthermore there is clearly a great deal of similarity between the affricates and the corresponding fricatives. In fact, we could have made many of the points concerning the general shapes of the tongue simply by reference to these data. We must note, however, some small but systematic differences between the affricates and the fricatives. To say that a sound is an affricate implies only that it involves a stop closure followed by a fricative. But in addition to the change in the tip or blade of the tongue that is required for making the stop preceding the fricative there is also a change in the positions of the body of tongue that is not necessarily implied by the traditional terms stop as opposed to fricative. The body of the tongue is often slightly higher during the stop closure than it is during the corresponding fricative.

If we overlook these minor discrepancies we can say that the place of articulation in the traditional sense is much the same in the affricates and the corresponding fricatives. In so far as the terms alveolar, retroflex, and alveolopalatal can be said to specify not just a place of articulation, but the shape of the tongue as a whole, then the same interpretation can be given to each of the categories for both fricatives and affricates. (But, as we have already noted, the same interpretation cannot be used for other manners of articulation; alveolar stops and nasals such as [t, d, n,] do not have the same tongue shape as the alveolar fricatives and affricates considered here.)

For the sake of completeness in our study of Pekingese fricatives, we must also mention [f] and [x]. The first of these sounds is a labiodental fricative with the lower lip in contact with the upper teeth as in many languages. The second is a weak velar fricative as shown in Figure 6. The three subjects have slightly different positions from one another. Subject A has the narrowest constriction in the front of the velar region. Subject B uses a distinctly further back position, including a narrowing in the upper part of the pharynx. Subject C is in between the other two, with the constriction being more nearly in what is traditionally described as the velar region.

The acoustic analysis of these fricatives will not be considered in detail here, but it is appropriate to discuss briefly how our articulatory findings can be correlated with some preliminary acoustic observations. The frequency of the turbulent noise associated with a fricative sound depends in part on the velocity of the air striking the sources of turbulence, and in part on the size of the cavity in front of the source of turbulence. Because of the first of these factors the mean frequency of the fricative noise in [s] is higher than that in [θ], and because of the second [c] is typically higher than [x]. The high velocity produced by the very narrow channel in [s], together with the more forward point of articulation and the smaller front cavity in comparison with [s], combine to produce the differences in frequency that distinguish this pair of sounds. Acoustic analyses (reported in more detail in Wu 1963) indicate that there is a wide spectral peak centered in the neighborhood of 6400 Hz for [s], with half power bandwidths at 5700 Hz and 10,000 Hz, whereas [s] has two lower and narrower peaks, one centered at about 2900 Hz, and the other at about 4500 Hz.

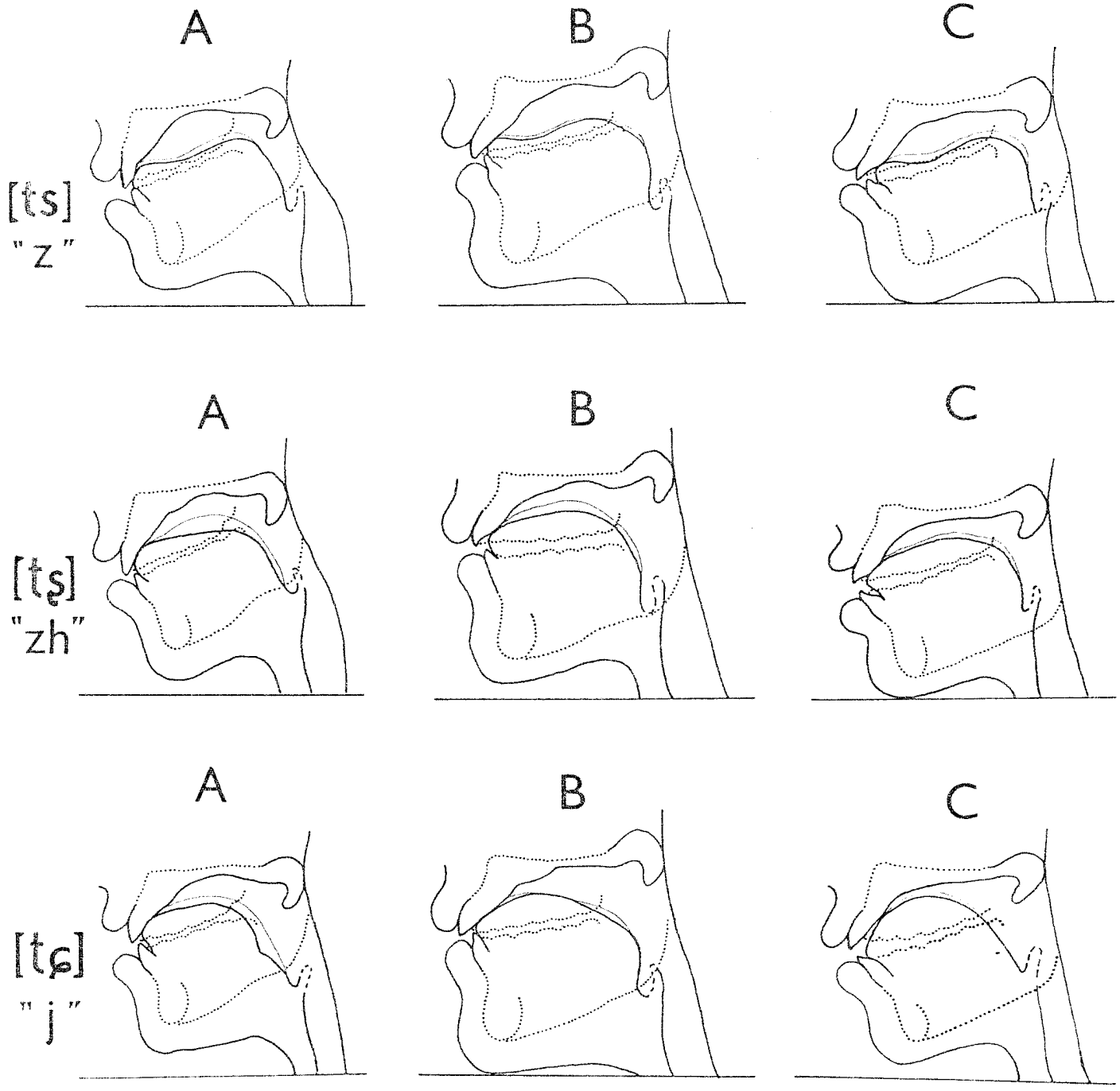


Figure 3. Tracings from x-rays of three speakers producing Pekingese voiceless unaspirated sibilant affricates.

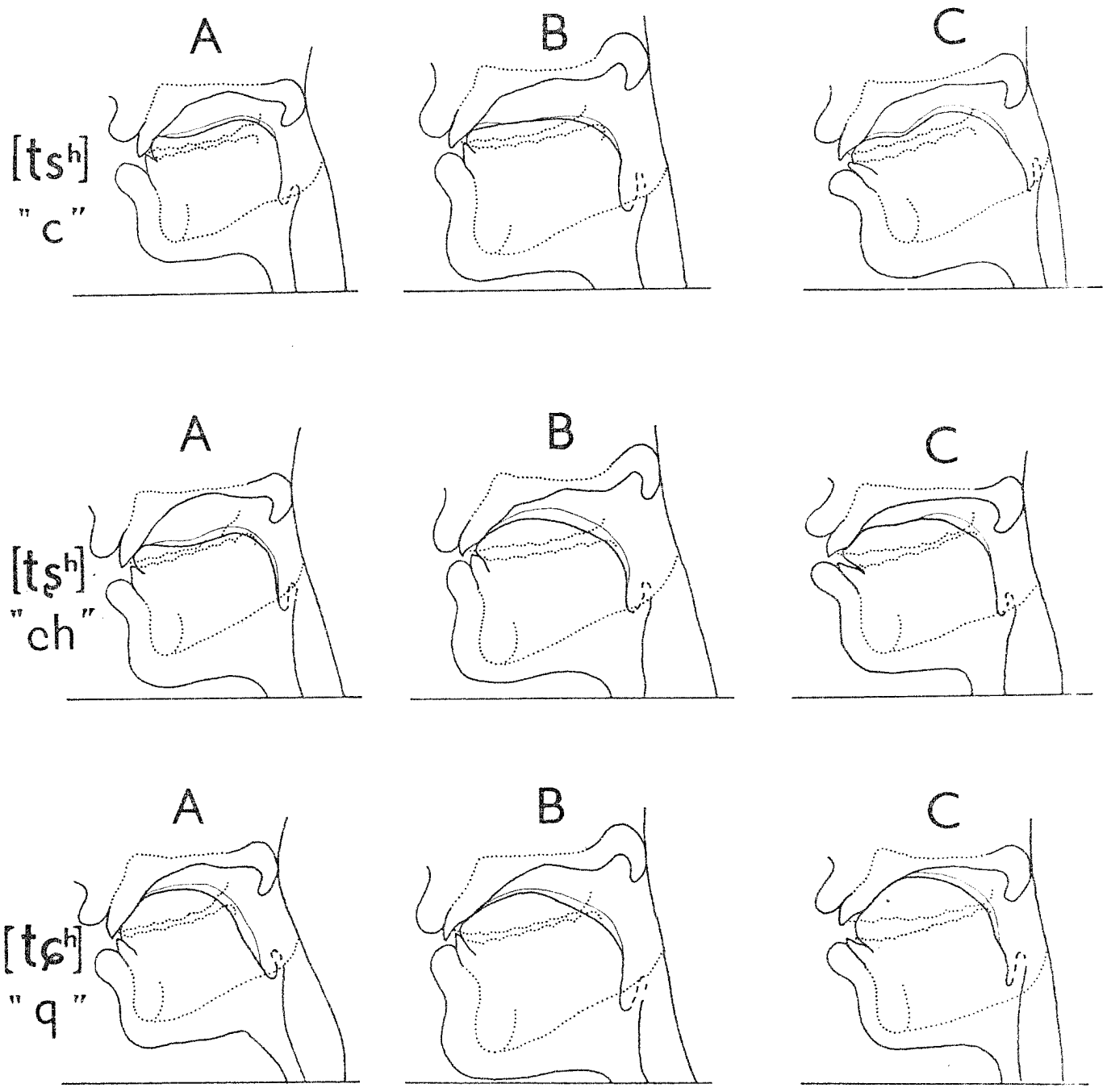


Figure 4. Tracings from x-rays of three speakers producing Pekingese aspirated sibilant affricates.

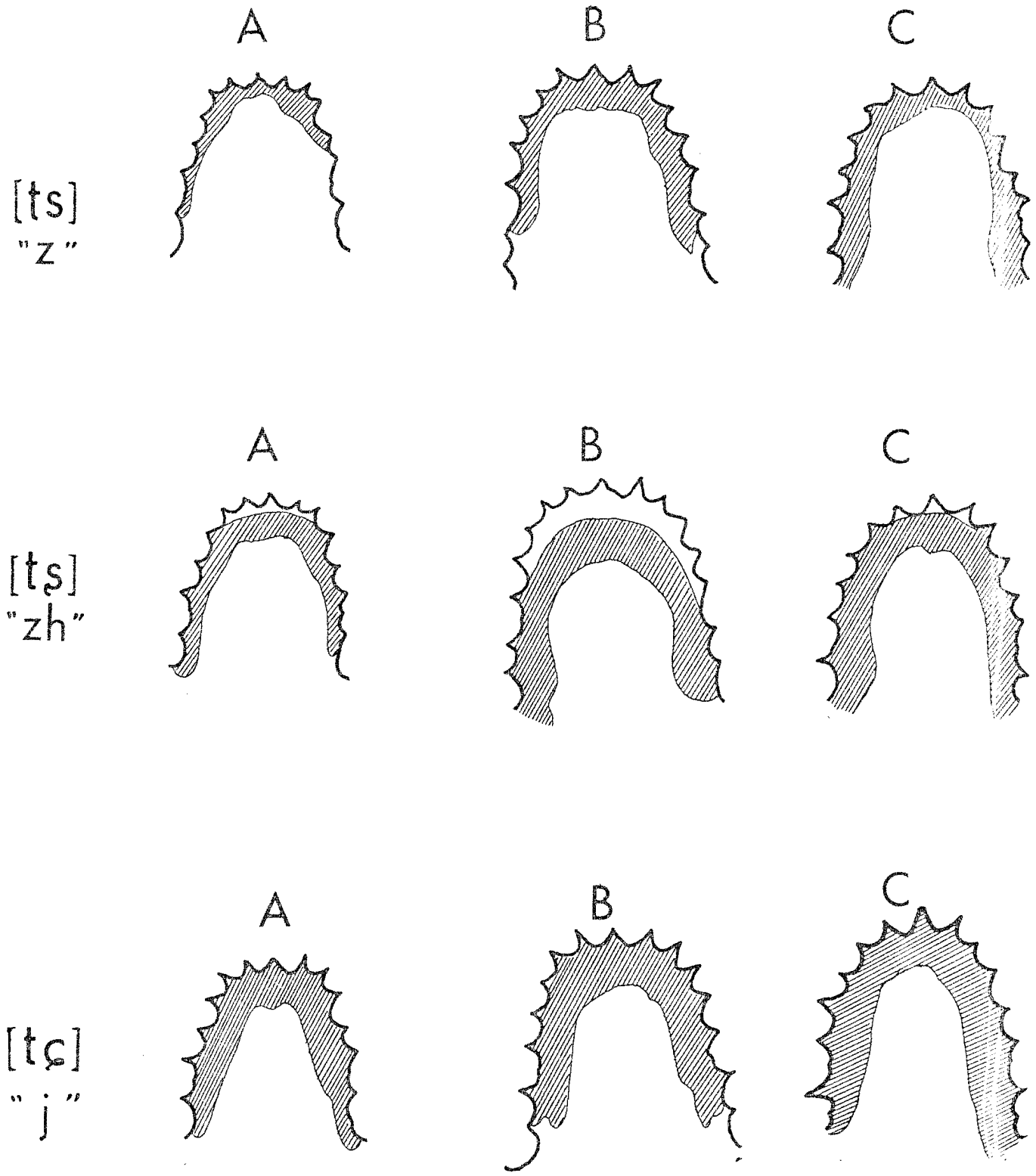


Figure 5. Palatograms of three speakers producing Pekingese sibilant affricates.

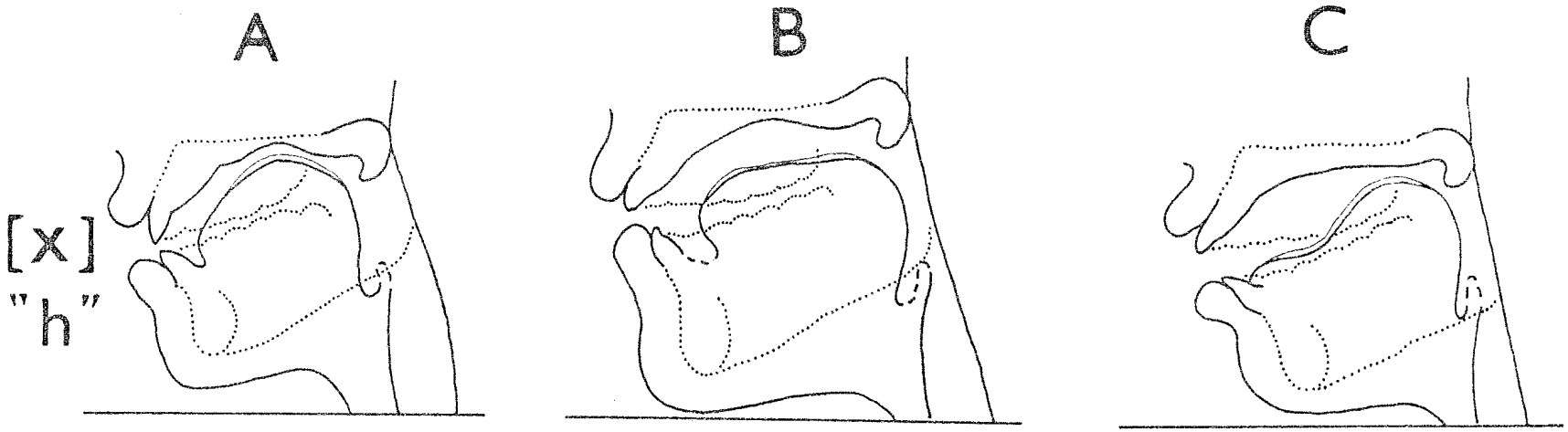


Figure 6. Tracings from x-rays of three speakers producing Pekingese non-sibilant fricatives.

General discussion

Throughout the presentation of the results we have been noting inadequacies in the notion place of articulation. But perhaps the basic question is really: Are there places of articulation? There are two ways of considering this question. Firstly we might ask, in the phonological description of a language, are there categories that group the sounds together in ways that can be thought of as corresponding to their places of articulation? The answer to this is, undoubtedly, yes. In nearly all languages, Chinese included, there are phonological rules that specify groups of sounds that can be said to be made at the same place of articulation.

The second way of considering the question is to ask whether there are, in a general phonetic sense, places of articulation that can be used in cross-linguistic descriptions of sounds, or even in the precise specification of the allophones that occur in a single language (for use, for example, in a synthesis by rule system). The answer to this is by no means clear, and leads us to consider the whole problem of how to relate phonological descriptions to observable phonetic data.

Phonology is concerned with describing (perhaps explaining) the patterns among the contrasting units in a language. Phoneticians want to describe (explain in terms of physiology or acoustics) the sounds that occur. The major problems in relating phonology and phonetics arise because of the different aims. They are made worse by the fact that an average language uses only about 31 contrasting segments, 23 of them being consonants (Maddieson, forthcoming). But the human vocal apparatus is capable of producing a vastly greater number of perceptually distinct sounds.

The usual practice of phonologists has been to give only a very approximate description of the sounds of a language, declaring that more detailed phonetic specifications could be made, but without showing exactly how this could be done. Many phonologists seem to feel that the phonetic details do not matter, and that precise accounts of the sounds of languages are not part of linguistics. But the details do matter to the phonetician who wants to teach people how to talk without a foreign accent, or who is trying to program a computer to reproduce the sounds of a language in a natural way. They also matter to any linguist who wants to make a complete, accurate, description of a language. Phonetic details are part of linguistics. Accordingly we are left with the problem of how to reconcile phonological and phonetic descriptions.

There are a number of phonologists who are very much concerned with phonetic detail. For example, Chomsky and Halle (1968) should certainly be absolved from the strictures against phonologists given above. They frequently point out small phonetic differences among languages, clearly regarding them as part of the phonology. The solution that they suggest is for the phonological description to specify "what the speaker of a language takes to be the phonetic properties of an utterance." But following this course would entail setting up a large number of phonological features to account for all the phonetic details of languages that speakers must know. There is no doubt that speakers of Pekingese take it that, for example, one of the phonetic properties of their alveolar fricative [s] is a deep hollowing of the tongue that is not found in their other fricative (or strident) sounds, or in their other alveolars. These speakers also consider their retroflex fricatives to have different phonetic properties from Telugu retroflex

fricatives; no speaker of Pekingese would make [ʂ] with the underside of the tip of the tongue. The raising of the tongue in affricates in comparison with fricatives also has to be taken into account by some feature, unless we can find a physiological explanation for its inevitability (meaning that every language always did it). In fact, for a linguistically adequate description, every particular aspect of tongue shape would have to be assigned to a particular feature, as Pike (1943) noted many years ago.

When faced with a similar problem involving the precise specification of the possible states of the glottis in different languages, Halle and Stevens (1971) opted to solve it by setting up a number of new features. They proposed that what speakers took to be the possible states of the glottis were whether it was [+ spread] or [- spread], [+ constricted] or [- constricted], [+ stiff] or [- stiff], and [+ slack] or [- slack]. Using these features they were able to suggest different specifications for several sounds that did not contrast within a language, but which nevertheless had different phonetic manifestations. For example, they specified English initial /p/ as in "pie" as [+ spread, - constricted, + stiff, - slack] and Korean so-called lax /p/ as [+ spread, - constricted, - stiff, - slack]. This same kind of detailed specification can be used for describing allophones within a language. Thus the stops in the English words "spy, sty, sky" can be differentiated from those in "pie, tie, kye" by calling them [- spread, - constricted, + stiff, - slack]. But there is a high price attached to this gain in phonetic accuracy. The more general categories [+ voice] and [- voice] have been given up. Few phonologists would be willing to do this. It is useful to consider English and many other languages, as having a contrast between [+ voice] and [- voice].

Following a similar approach in describing places of articulation would lead to similar difficulties. We could create more categories by dividing the columns in a consonant chart. This is the course the IPA has followed over the years since its founding, although clearly with some reluctance. The category alveolopalatal, which we have been using to describe some of the sounds of Pekingese, does not currently have the status of a column heading. It is listed (IPA 1979) among the miscellaneous items below the chart as a possibility for use in describing fricatives. It is not available for describing stops or any other sounds. If we were to add alveolopalatal as a separate column heading (as it was in earlier editions, IPA 1949), plus another column so that we could distinguish the retroflex sounds of Pekingese from those of Telugu, yet other categories (which may involve additional rows such as "Grooved") to allow for the hollowing of the tongue in some sounds but not others, and so on, we would soon have a very unwieldy set of features. It would be as useless to phonologists as are specifications such as [+ spread], [- constricted], [+ stiff], [- slack]. The IPA (1949) obviously recognizes this in its comments such as: "The more familiar letters ʃ ,ʒ may be used to denote the sounds ɕ,ʑ in languages like Pekingese, which contain these sounds and do not contain the more usual varieties of .,.."

We may be able to handle this problem with the aid of cover features, as suggested by Ladefoged (1972) and Vennemann and Ladefoged (1973). In the case of the glottal features, Stevens (1983) has proposed something very similar. He has listed a set of features that includes Voiced as well as the four glottal features given above. Although Stevens might not put it in this way, this makes Voiced a cover feature definable entirely in terms of specific combinations of the other four features. The feature Voiced then becomes available for phonological rules, leaving the other four features available for detailed phonetic specification of the states of the glottis.

We do not know how this notion should be applied to places of articulation. We could set up cover features such as Coronal (or Lingual) for articulations in the dental, alveolar, and a little bit further back region, and Dorsal for the uvular, velar, palatal, and a little bit further forward region. But it is not apparent how such cover features could be defined in terms of more precise phonetic features. It might be possible to do this in terms of acoustic properties, but it seems to us that the acoustic correlates of places of articulation for stops (both exploded and unexploded), nasals, and fricatives are not likely to have much in common; so acoustics is probably not the answer to giving suitable definitions of places of articulation.

An alternative possibility is that we should consider the phonological categories as specifying only approximate phonetic properties of the sounds. Doing this would enable us to equate sounds in different languages; but again it is not at all clear how it could be done. For example, there are five Pekingese voiceless fricatives /f,s,ʂ,ɕ,x/ and four English sounds /f,θ,s,ʃ/ in the same category. But how can we equate them? English /f/ goes with Pekingese /f/ all right, but it would be very odd to put English /θ/ in the same category as Pekingese /s/. But even when we have left English /θ/ unpaired, so that we can put English /s/ with Pekingese /s/, our troubles are not over. Which is the most appropriate category for English /ʃ/? Is it the same as Pekingese /ʂ/ or /ɕ/? As we have seen, the IPA answered this by saying that they are all different; and we can do no better. If we are trying to associate each feature with a specific phonetic property, we will have to say that English /ʃ/ has a feature specification that is different from any of the Pekingese fricatives. We cannot see what phonetic property English /ʃ/ has in common with one of the Pekingese fricatives /ʂ,ɕ/, which they do not also have with each other. We would add that, despite the IPA, we can see no motivation for classing Tamil [ʂ] with Putonghua [ʂ]. They are as different as Pekingese [ʂ] and [ɕ]. Within English, Pekingese, and Tamil there seem to be <u>eight</u> phonetically identifiable places for fricative sounds: [f,θ,s,ʃ,ɕ, (Pekingese) ʂ, (Tamil) ʂ,x]. Given other languages there could be even more. As we noted earlier, we are not at all sure that the Polish sound that the IPA symbolizes [ɕ] is the same as the Pekingese sound which is also represented by this symbol. Is there a specific number of possible fricative sounds?

So we are really no further forward. We have not escaped the dilemma described in the preceding paragraphs. There is no way of non-arbitrarily assigning sounds to approximately specified places of articulations. The feature system must be rich enough to allow for the specification of all the distinct places of articulation (up to seven contrasting stops occur in Yanuwa (Ladefoged 1983). It must also provide a non-arbitrary way of relating all the sounds of a language to some specific phonetic properties. But if it does all this adequately it will be inconvenient for use in describing the phonological patterns within a language.

There is no simple way out of this problem. When we are considering ourselves as phonologists, we will continue to describe the patterns among the contrasting sounds. As phoneticians, we will continue to describe the actual sounds that occur. Phonologists must behave as if there were distinct places of articulation, grouping sounds together in ways that are appropriate for the particular language being described. Meanwhile phoneticians will have to go on doing their best to specify the sounds of each language in general anatomical and acoustic terms. They will not be able to allocate consonants to a small number of cells on a chart, just as they cannot describe vowel qualities or tonal contrasts

in terms of a set of distinct phonetic properties. Nor is it simply a matter of not being able to define the boundaries between adjacent places. Languages divide up the continuum of possible places of articulation in different ways, much in the same way as they divide up the tone and vowel spaces in different ways. There are, of course, favored regions that occur in many languages. But the phones that are grouped together phonologically in one language will not be the same as those that are grouped together in another. Nobody imagines that the vowel and tone spaces are divided into specific sets of categories. Why should we imagine that there are discrete place of articulation?

References

Balasubramaniam, T. (1972). The Phonetics of colloquial Tamil. Ph.D. Thesis, University of Edinburgh.

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

Halle, M. and Stevens, K.N. (1971). A note on laryngeal features. Quarterly Progress Report of the Research Laboratory of Electronics 101, 198-213. Massachusetts Institute of Technology.

Hammarstrom, G. (1957). "Uber die Anwendungsmoglichkeiten der Palatographie."Zeitschrift fur Phonetik und allgemeine Sprachwissenschaft. 10.4, 323-336.

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

Ladefoged, P. (1972). Phonological features and their phonetic correlates. Journal of the International Phonetic Association 2, 2-12.

Ladefoged, P. (1983). Cross-linguistic studies of speech production. Mechanisms of speech production (ed. by P. Macneilage) New York: Springer-Verlag. 177-88.

Ladefoged, P. and Bhaskararao, P. (1983). Non-quantal aspects of consonant production. Journal of Phonetics. 11,291-302.

Maddieson, I. (forthcoming) Patterns of sounds. Cambridge: Cambridge University Press.

Pike, K. (1943). Phonetics. Ann Arbor: University of Michigan Press.

Puppel, S. (1977). A Handbook of Polish Pronunciation. Warsaw: Państwowe Wydawnictwo Naukowe.

Stevens, K. (1983). Personal communication at the Symposium on Invariance and Variability in Speech Processes MIT

Vennemann, T. and Ladefoged, P. (1973). Phonetic features and phonological features. Lingua 32, 61-74.

Wu, Z. (ed) (1963). Phonetics of Putonghua Consonants, Phonetic Laboratory, Institute of Linguistics Academia Simica. (Unpublished manuscript).

Is there a valid distinction between voiceless lateral approximants and fricatives?

Ian Maddieson and Karen Emmorey

Paper presented at the 10th International Congress of Phonetic Sciences
Utrecht, August 1983.

Many schemes of phonetic classification propose that there is only one manner of articulation for voiceless laterals, namely that they are fricative. Pike, in his classic book Phonetics, expressed this view as follows: "Laterals...upon becoming voiceless narrow the opening sufficiently to get local friction" (1943: 72). Catford (1977: 132) also implies that phonemic voiceless laterals are invariably fricatives. And many practising field linguists seem to feel that once they have said a lateral is voiceless, they have said all that is necessary. For other phoneticians, voiceless lateral fricatives and voiceless lateral approximants are distinct types of sounds. This view was taken by Ladefoged in Preliminaries to Linguistic Phonetics, but he adds that the difference is not used contrastively in languages. He says, "the contrast between voiceless lateral approximants and lateral fricatives occurs only among voiced sounds. I do not know of any language that distinguishes between voiceless lateral fricatives and approximants, although many languages...have one or the other of these sounds" (1971: 53) (emphasis added). More recently Maddieson (1980; forthcoming) has also commented on the lack of examples of languages reported to have voiceless laterals of both these types.

But are these two types of sounds really different from each other? If there is no reliable way of distinguishing them, it follows that no language would use them contrastively, and the claims of Ladefoged and Maddieson about the avoidance of such contrast would be vacuous. We would be dealing with a difference in terminology and transcriptional practise that has no real phonetic basis, but is instead a reflection of different traditions and habits among linguists. In general, linguists working on languages of Africa, the Americas, and Europe favor terms and transcriptions that imply that the voiceless laterals in these languages are fricative, whereas those working on Asian languages favor terms and transcriptions that imply that the voiceless laterals are approximants. However, if this is not merely a reflection of linguists' habits, then it would imply that there are important areal/genetic assymetries in the distribution of the two sound types.

This paper sets out to establish that there is a phonetic distinction between two types of voiceless laterals and moreover that the phonetic difference has important phonological consequences as well.

Speakers of several languages with voiceless laterals were recorded saying words with these sounds in initial position. Among those languages reported to have voiceless lateral fricatives we recorded nine speakers of Navaho, three speakers of Zulu, and eight speakers of Taishan Chinese, as well as examining several individual speakers of other languages. Among those languages reported to have voiceless lateral approximants, we recorded four speakers of Burmese, and three speakers of Tibetan (one of whom speaks the Sherpa dialect). The majority of these speakers were recorded in the UCLA Phonetics laboratory under the same conditions. Where possible, words with a low central vowel after the lateral were used. As we will see, these five languages form two pairs, the fricative

languages Navaho and Zulu vs. the approximant languages Burmese and Tibetan, with Taishan occupying an intermediate position.

Three types of measures were made, measures of duration, amplitude and spectral shape. The procedures used and the results obtained with respect to each of these will be presented in the order just given. Figure 1 shows how durations were measured from a display of the digitized waveform on a computer screen. The duration of the noisy portion of each lateral was measured from the onset of the lateral to the onset of voicing. The duration of any voiced lateral portion before the vowel was also measured. As in the particular tokens from Burmese and Zulu shown in Figure 1, this voiced portion is typically longer with the voiceless approximants than with the voiceless fricative laterals, and the voiceless noisy portion is correspondingly shorter. Mean durations of these two portions for all tokens from all speakers of each language are shown in Table 1. In addition, the duration of the voiced portion as a percentage of the total duration is given.

	noise	voice	proportion of voicing	
Zulu	232	17	6.8%	
Navaho	195	15	7.1%	
			----------	significant difference
Taishan	150	21	13.3%	
Tibetan	111	18	14.8%	
			----------	significant difference
Burmese	136	55	29.2%	

Table 1: Duration measures

The languages with the longest durations of noise are those with voiceless fricative laterals. Burmese has a markedly longer voiced portion in its voiceless laterals than the other languages. The best measure for comparing the durations is the percentage measure, since this normalizes for possible speech rate differences. This measure divides the languages into three groups, as shown in the table (significance was computed using Tukey's Studentized range test with a criterion of .01). Note that Taishan and Tibetan are not distinguished, although Zulu and Navaho are significantly different from Tibetan and Burmese.

The relative amplitude of the voiceless lateral with respect to the amplitude of a following low central vowel was measured from amplitude envelope displays like those in Figure 2, produced by the Kay digital spectrograph. In general, as in the particular tokens from Navaho and Tibetan shown in Figure 2, the amplitude in the voiceless approximants is less than in the following vowel, whereas in the voiceless fricatives, the amplitudes of the lateral is closer to that of the vowel. Table 2 shows the mean amplitude differences between lateral and vowel in all the tokens of each language. Units of measurement are arbitrary but constant across languages.

	mean	
Zulu	0.24	
Navaho	0.44	
	---------	significant difference
Taishan	1.46	
Burmese	1.58	
	---------	significant difference
Tibetan	3.53	

Table 2: Lateral/Vowel Amplitude Difference

FIGURE 1

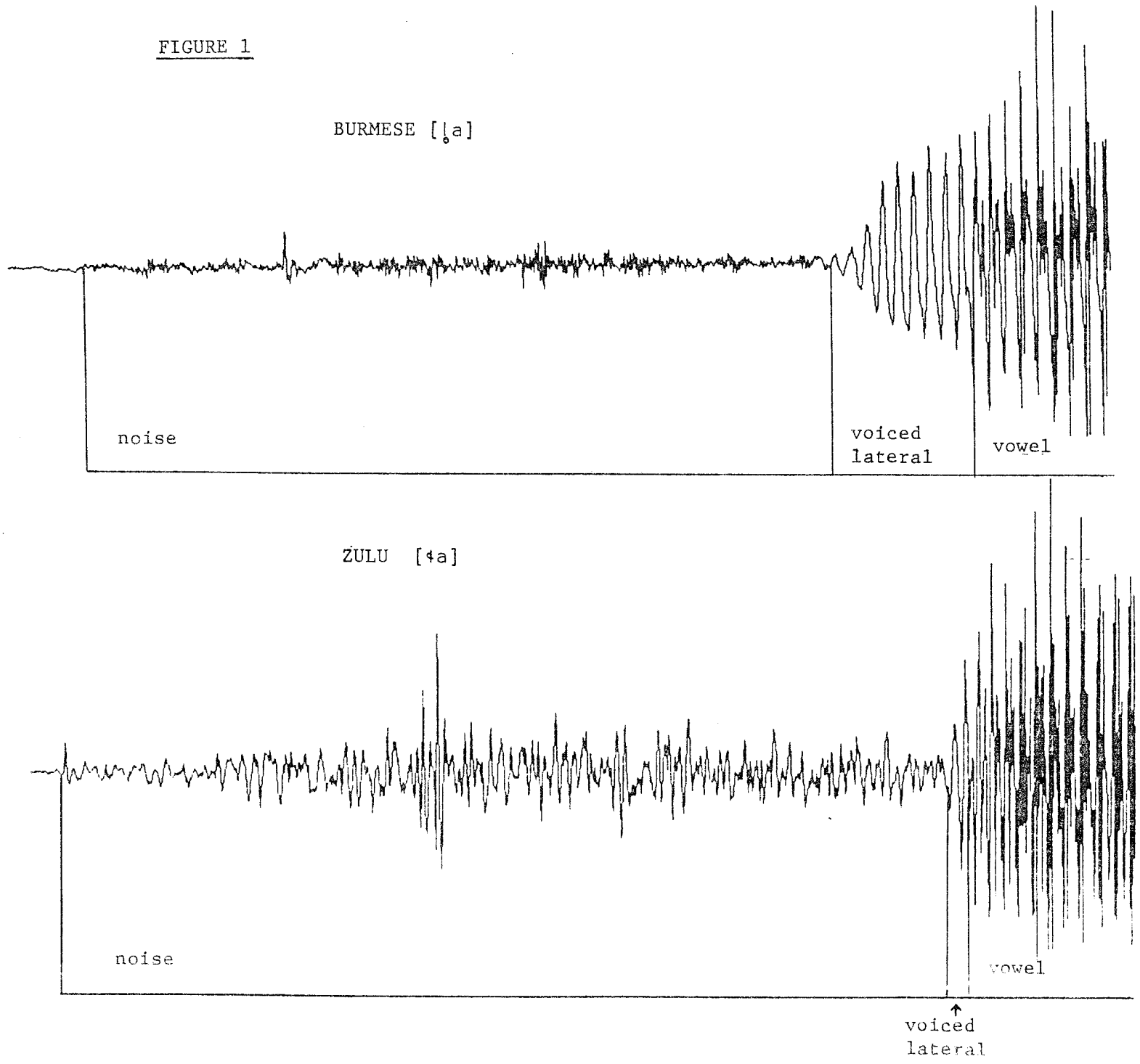
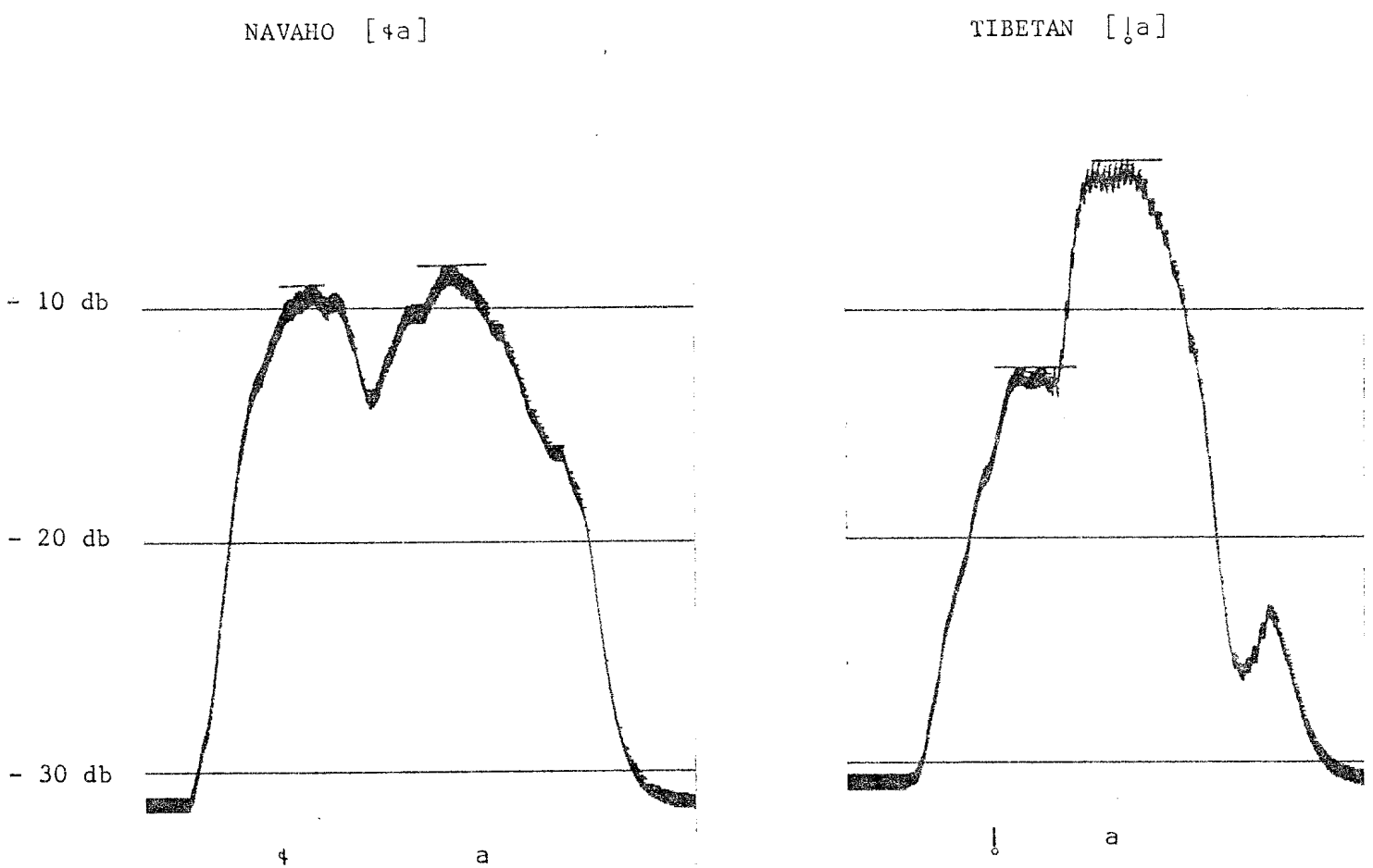

BURMESE [l̥a]
noise
voiced
lateral
vowel
ZULU [ɬa]
noise
vowel
voiced
lateral

FIGURE 2



On this measure Zulu and Navaho are again significantly different from the other three languages. Additionally, Tibetan is significantly different from all the other languages. Taishan is not distinguished from Burmese.

The third property of the voiceless laterals examined was the spectrum between 100 and 9000 Hz. An FFT analysis was made of the central portion of the noise in the laterals (all tokens analyzed were word-initial preceding /a/). This spectrum was then convolved with a filter designed to approximate the functioning of the human auditory system and the output was expressed as values of 20 critical bands. (For more information on the analysis system used, see Nartey 1982: 21-33.) The critical band values were then analyzed by means of canonical discriminant analysis, using language as the classification variable. Although four canonical variables are statistically significant, only the first has an interpretation which relates to the fricative/approximant distinction. Figure 3 shows a plot of the first two canonical variables. Each letter represents the position in this two dimensional space of one token from the language indicated. As may be seen, the Zulu and Navaho tokens cluster to the right of the diagram with similar values on the first canonical variable. Burmese and Tibetan are distributed over the center and left. Taishan is in the center right area. This plot shows again that there are differences between the two types of voiceless laterals with Taishan tending to be intermediate. Note that on this measure Taishan is more similar to Zulu and Navaho than it is on the duration and amplitude measures. In the other dimension shown, it can be seen that the two characteristically "fricative" languages, Zulu and Navaho, are the most distinct from each other. We therefore assume that this (and, likewise, the other higher canonical variables) must be unrelated to the distinction we are investigating. We will attempt an interpretation of only the first canonical variable.

Examination of the canonical coefficients indicates that the major relevant differences are to be found in critical bands 7, 8, 10, 15, 16, 17, 18, 19 and 20, since these have the highest (positive or negative) coefficients. Mean amplitudes in these bands for each language in the critical band spectra were compared. Simplifying the result somewhat, we find that the fricative laterals have high energy in a region of the spectrum from 3150-6400 Hz. (bands 16-19), whereas the approximant laterals have high energy in the band below this region (band 15, 2700-3150 Hz.). All the languages show relatively low energy in band 10 (1270-1480 Hz.), but it is particularly low in Tibetan, the language with the highest negative coefficient on the canonical variable. In band 8, Zulu, Navaho and Burmese all have low values, whereas Tibetan and Taishan have higher values. It is less obvious how band 7 (770-920 Hz.) contributes to the discrimination. Most obviously, the fricative laterals have a broad high frequency peak, while voiceless approximant laterals reach their peak at a lower frequency.

In summary, the evidence from our three measures shows that the two types of voiceless laterals can be distinguished and that languages such as Navaho and Zulu on the one hand and Tibetan and Burmese on the other provide good archetypes of the voiceless fricative and approximant types respectively. Furthermore, the nature of these distinctions correlates well with the labels given; that is, as fricatives vs. approximants. The fricatives tend to have later onset of voicing, relatively greater noise amplitude and greater energy at high frequency than the approximants. These are all reasonable effects to find as results of a difference in aperture. However, as with many other phonetic variables, there is gradience in this distinction, and the languages differ from each other by degrees and not by completely categorical divisions. In particular, Taishan presents itself as an

intermediate case. On all three measures it is closer to Zulu and Navaho than Tibetan and Burmese are. It should perhaps be noted that Taishan is the only one of the languages compared in which the voiceless lateral is dental rather than alveolar in articulation. Furthermore, the segment [l̪̊] in Taishan varies with [θ] in the speech of several of our speakers.

It should be recognized that despite our ability to document a difference, the voiceless lateral fricative/approximant distinction remains a subtle one. It has been argued that the absence of languages which contrast these sounds can be explained by the lack of sufficient saliency for the differences between them (Maddieson, to appear 1984). It is therefore perhaps surprising to find that there are notable differences in the phonological behavior of the two types.

Several phonological differences were observed in a survey of some 60 or so languages with voiceless laterals. These involve phonotactics, allophonic variation, and co-occurrence with other laterals in the inventory. The most striking of them may be briefly stated as follows:

(1) Voiceless lateral approximants are restricted to syllable initial position; fricatives aren't.
(2) Voiceless lateral fricatives may have affricate allophones; approximants don't.
(3) Voiceless lateral approximants always occur together with a voiced lateral approximant in the inventory; voiceless lateral fricatives may occur without a voiced lateral.

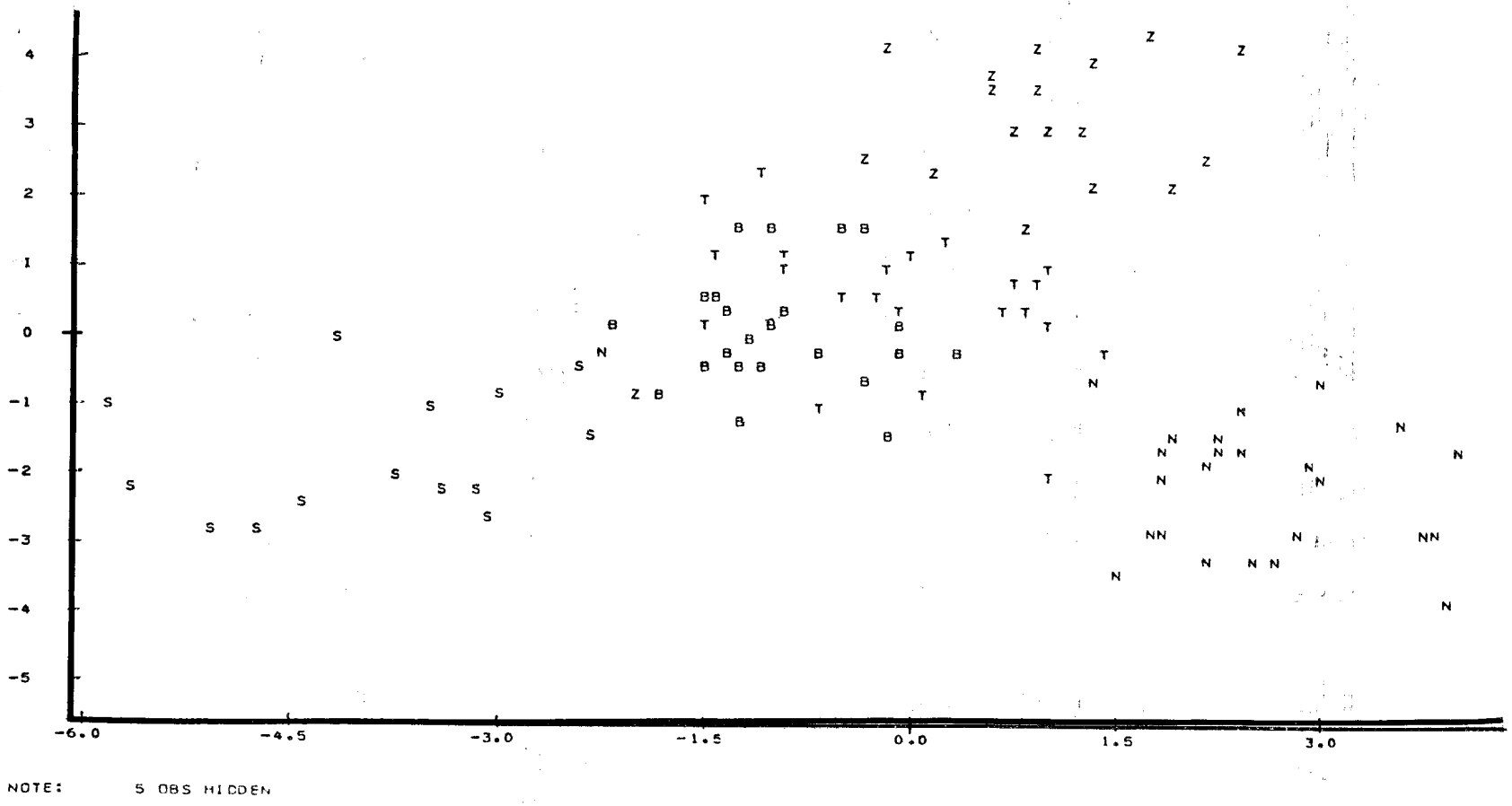
The restriction of /l̥/ to initial position is only of serious value in establishing a difference between /ɬ/ and /l̥/ when there are not more general restrictions on final consonants. Thus, although /l̥/ is restricted to initial position in Yao, Burmese, Iai and Tibetan, so also is /l/. But in Klamath, S. Khmu?, Kuy, Sedang and Irish (in those dialects in which mutation of /sl/ and /ʃl/ is the only source of /l̥/), /l/ has greater freedom of occurrence than /l̥/, including in final position. As for the possibility of affricate allophones of /ɬ/, these occur in intervocalic positions in Alabama and Nez Perce, after nasals in Totonac and Zulu, finally in Tolowa and initially in long syllables in Hupa. In most of these cases, the affricate is an optional variant. Thirdly, Tlingit, Nootka, Puget Sound Salish, Chukchi and Kabardian are among those languages with voiceless lateral fricatives but no voiced lateral approximants. Proto-Chadic (Newman 1977) is reconstructed with /l̥/ but no /l/ and some of the modern Chadic languages retain this feature (e.g. Bade and Ngizim, at least with respect to native vocabulary). Thus the phonetic difference between voiceless lateral fricatives and approximants is a significant one in phonological terms. There are considerably more languages with fricative lateral phonemes than with voiceless lateral approximants, suggesting that the latter are diachronically less stable (e.g. the formerly distinct /l/ and /l̥/ in Proto-Tai have merged as /l/, Li 1977) or have fewer possible sources. This may point to another difference between them. In any case, to collapse these two types of laterals into a single class is to miss an important difference.

Acknowledgments

This research was funded by a grant from the National Science Foundation for research into phonetic differences within and between languages. Grateful thanks are due to Jonas Nartey and Peter Ladefoged for sharing data they had collected and to Eric Zee for assistance in contacting and recording speakers of Taishan Chinese.

Figure 3
Plot showing locations of individual tokens of voiceless laterals in a two-dimensional space defined by coefficient values of the first two canonical variables in the canonical discriminant analysis of the critical band spectra.

S = Tibetan, B = Burmese, T = Taishan, Z = Zulu, N = Navaho



References

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

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

Maddieson, I. 1980. A survey of liquids. UCLA Working Papers in Phonetics 50: 93-112.

--------- to appear (1984). Patterns of Sounds. Cambridge University Press, London.

Li, F-K. 1977. A Handbook of Comparative Tai. University of Hawaii Press, Honolulu.

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

Newman, P. 1977. Chadic Classification and Reconstructions. Afrosasiatic Linguistics 5.1.

Pike, K.L. 1943. Phonetics. University of Michigan Press, Ann Arbor.

Phonetic Cues to Syllabification

Ian Maddieson

Introduction

Ladefoged (1982:219) states simply that "there is no agreed phonetic definition of a syllable". That such a definition is lacking can be readily seen in any reading of current phonetic literature. Earlier optimism over defining the syllable (see Pike 1943, Stetson 1951, inter alia) has largely given way to pessimism. Yet despite the difficulty of defining it, the syllable has been given a major role in recent developments in phonological theory (e.g. by Kahn 1976, Kiparsky 1979, Selkirk 1980, Steriade 1982, Cairns and Feinstein 1982, Clements and Keyser 1983, inter alia). Views differ as to how complex the internal structure of the syllable is, for example over whether a syllable node dominates higher order elements with their own constituent structure such as onset and rhyme, dominates C and V elements, or directly dominates segments (i.e. feature matrices). However, all accounts essentially agree that in some way segment-like elements are grouped into syllables.

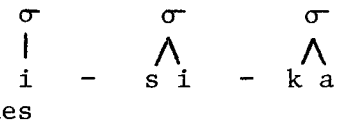
Moreover, there may be rules which change the membership of a segment from one syllable to another (resyllabification rules). For example, Harris (1983) states a common observation about Spanish as follows: "in casual speech a word-final consonant syllabifies with the initial vowel of the following word" (p. 43). He formulates a rule which reassigns a consonant before a word boundary and a vowel to an onset rather than a rhyme, and exemplifies the effect of the rule with the sentence:

Los otros estaban en el avión.

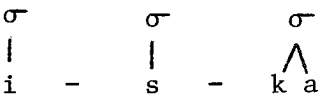
After this rule applies this sentence is syllabified as follows (a syllable boundary is represented by a period):

Lo.so.tro.ses.ta.ba.ne.ne.la.vión

Elsewhere, Marlett and Stemberger (1983) argue that resyllabification of a somewhat different sort applies in Seri after vowel deletion in certain forms with prefixes. The prefixes have the form consonant + /i/, as in the irrealis /si-/. When a consonant follows this prefix, the vowel /i/ is dropped so that



becomes



Since a syllable containing only /s/ is not well-formed, a rule of resyllabification applies which attaches /s/ to the onset of the following syllable, giving /i.ska/.

With the ability to define a syllable phonetically in doubt, questions obviously arise concerning the basis on which selection between plausible alternative syllabifications is made. These questions apply as much to any initial (lexical) assignment to syllables as to cases where resyllabification is posited across word boundaries, as in the Spanish example above. Likewise, in the Seri example the syllabification of /s/ with /ka/ is obviously only one of two

potential ways that the faulty syllabic structure could have been remedied. The output of resyllabification could have been /is.ka/, with the syllable boundary between the consonants as in the Spanish example /es.ta.ban/.

Determining syllabification

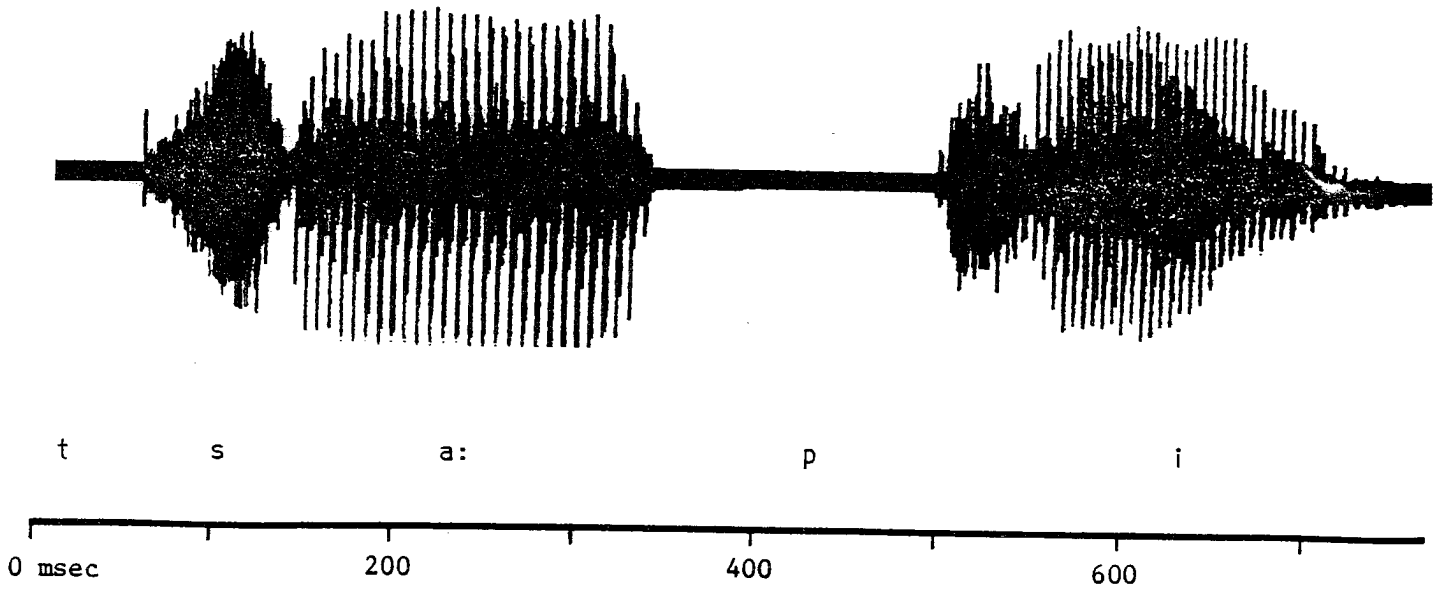
When syllabification is at issue what is the basis on which linguists have determined that their view in any given case is the correct one? There are sometimes formal arguments that can be made to justify particular syllabifications, e.g. those related to input to reduplication rules advanced in Steriade (1982). These seem more likely to relate to initial syllabification, leaving surface syllabification to be determined from other kinds of evidence. If it is accepted that languages differ in their surface syllabification[1], it is reasonable to assume that there must be some indications of the differences in syllable structure in the phonetic string. Note that the lack of a phonetic definition of the syllable does not prevent the recognition of phonetic markers of syllable constituency.[2] Their presence would enable a resolution to be made in situations where alternative syllabifications might be posited.

Of course, in many languages there are extrinsic allophonic rules which select allophones based on their position in the syllable. A well-known example is the difference between syllable-initial and syllable-final /l/ in both British and American English. Acoustic data on this phenomenon in American English is provided in Lehiste (1964). Since it is syllable-based, this allophonic difference is capable of providing evidence for constituency of syllables in potentially ambiguous cases. An example is "holy" vs. "holey" (i.e. "hole" + adjectival suffix "-y"). The word "holy" is syllabified [hou.li] with the syllable-initial ("clear") allophone of /l/. In the monosyllable "hole" a syllable-final allophone of /l/ occurs and, in my speech as well as that of many other speakers of British English, a special allophone of the preceding vocalic element that occurs only before a tautosyllabic lateral also occurs. That both of these features occur in the derived form "holey" provides evidence that the syllabification of this word retains the lateral as a constituent of the first syllable (cf. Faure 1972). Tokens of "holy" and "holey" showing these syllable bound properties from my speech are given in Figure 1.

Such an allophonic difference in laterals (and in vowels preceding laterals) is a particular fact about my dialect of English and is not general across languages.[3] Many languages lack such salient cues for syllable constituency in their allophonic rules. And in the languages which do show them they are not present in all segment types. It follows that if there are cues to syllabic constituency in these other situations, they must be more subtle ones. Linguists (not to mention native speakers) may well be responding to these cues when they make judgments about syllabic constituency in their data.

An explicit appeal to these more subtle cues can be made in the attempt to determine syllabic constituency in ambiguous circumstances. For example, Maddieson (1983) claims that most word-initial consonant sequences in the Chadic language Bura are resyllabified when a vowel precedes. Thus the first element of the sequence becomes a coda to the syllable containing that vowel and the syllable boundary falls between the elements of the sequence. For example, the verb /bda/ when preceded by the person/aspect marker /tsa:/ is syllabified as /tsa:b.da/, and similarly with other sequences. Part of the evidence for this view is that the vowel preceding one of these sequences tends to be shorter than that before a single word-initial consonant. Figure 2 shows waveforms of

(a) before a single consonant:



(b) before a consonant sequence:

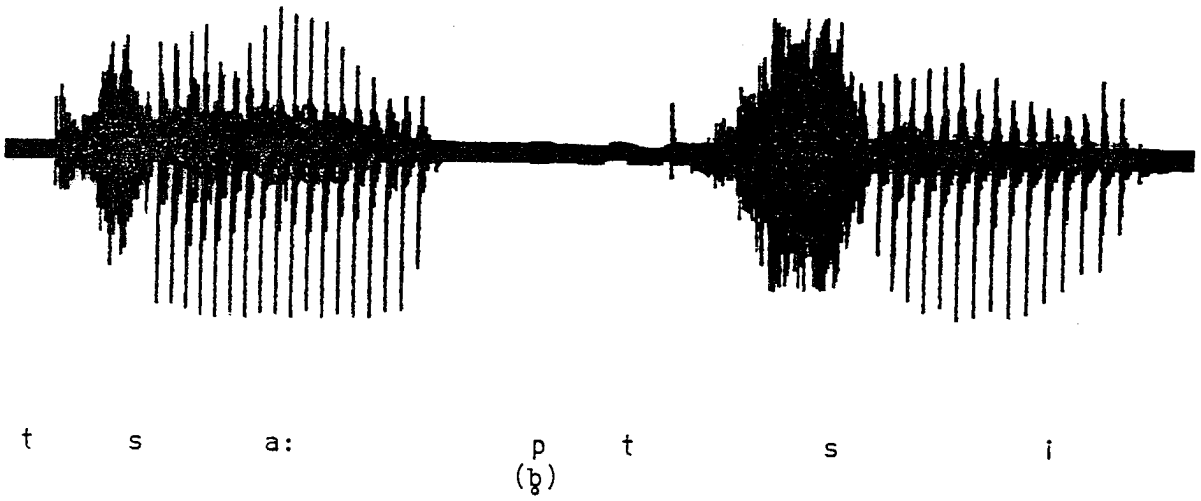


Figure 2. Vowel duration in the morpheme /tsa:/ before word-initial (a) single consonant, and (b) consonant sequence in Bura (from Maddieson 1983).

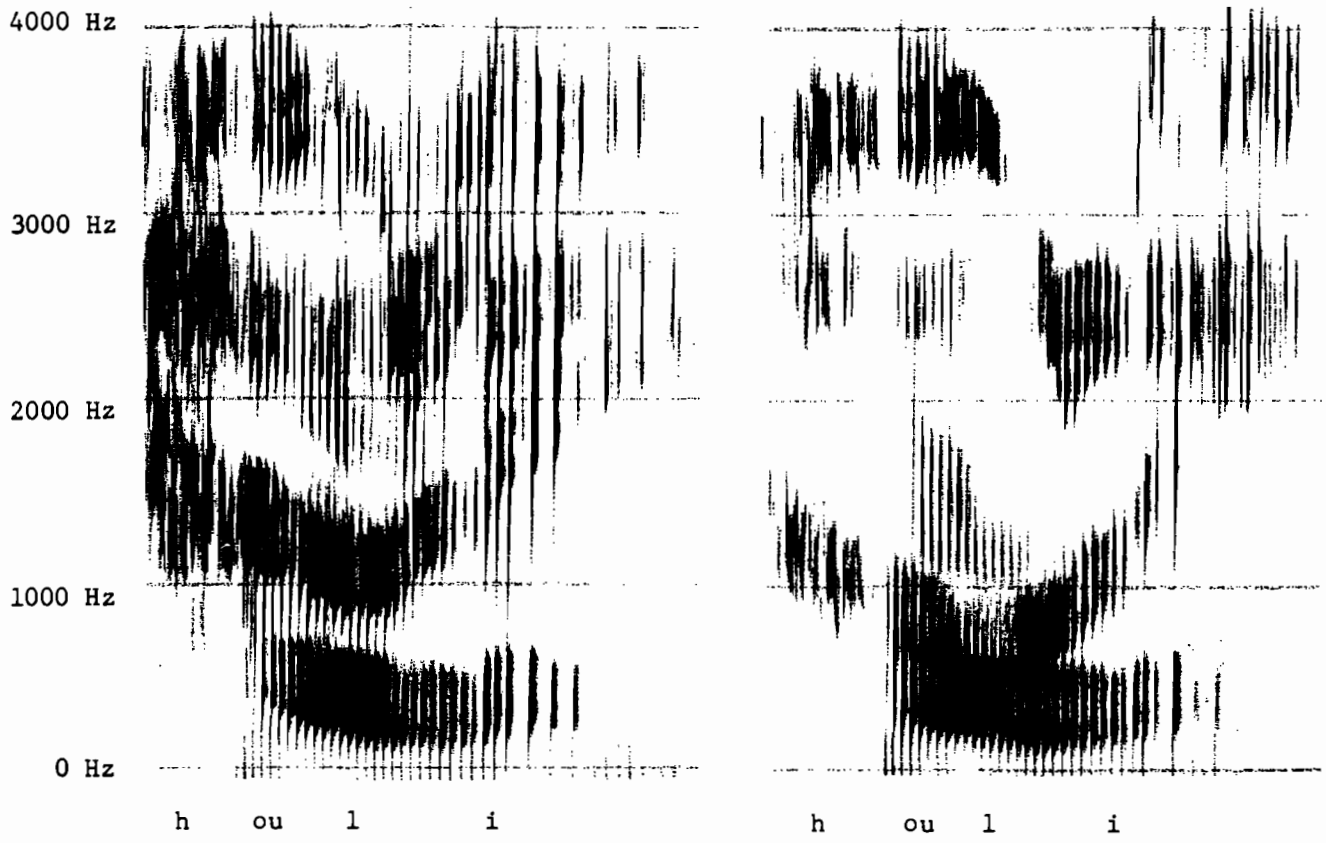


Figure 2. Syllable dependent allophones of /l/ in the near-homophones holy (left) and holey (right) in the author's speech. Note F_2 above 1000 Hz in holy but below 1000 Hz in the vowel /ou/ and the lateral in holey.

utterances containing /tsa:/ before [ptsi] and [pi]. A substantially shorter vowel can be seen before the sequence consisting of [p] preceding the affricate /ts/ (Maddieson 1983 argues that this [p] is actually an underlying /b/) than before the single consonant /p/ in the form /pi/.[4] As will be shown below, vowel shortening in closed syllables is a relatively common phenomenon across languages. Its occurrence in this Bura example thus provides some objective support for the intuitive feeling that the syllable boundary falls where it is shown in /tsa:b.tsi/.

However the appeal to vowel shortening in Bura carries no weight unless the phenomenon of vowel shortening in closed syllables is in fact a general cross-linguistic one. Since there are no Bura words which contain unambiguous syllable-closing obstruents (e.g. word-final stops), the argument cannot be extended to the between-word cases on the basis of language-internal evidence from within-word cases. Unless closed-syllable vowel shortening can be shown to be quite widely found in other languages the Bura data are unconvincing evidence for any particular syllabification, since the mere fact of a difference could reflect a language-particular rule of vowel shortening under some other circumstances.

The topic of universal bases for recognition of syllable constituency seems to have been rather neglected since an early and unconvincing experiment on formant transitions with the Haskins pattern-playback synthesizer by Malmberg (1955 [1967]). Although the vowel shortening referred to above has been mentioned as common before (e.g. by Jones 1950, Abercrombie 1967), there does not seem to be any study which has explicitly shown that it tends towards universality and hence has value as evidence. The remainder of this paper is dedicated to showing that vowel shortening associated with syllable structure is widely found.[5] For convenience the phenomena being investigated will be referred to under the name Closed Syllable Vowel Shortening (CSVS).

Vowel duration

Of course, many other factors beside syllable constituency affect the duration of a vowel including lexical vowel quantity and other inherent properties of the vowel itself, as well as various "suprasegmental" factors (stress, tone, intonation, etc.), and contextual effects such as the nature of surrounding segments and position in units such as the word, sentence, etc. However, when these factors are controlled for, many languages prove to have a vowel duration difference that relates to the syllabification of the following consonant. It will also be shown below that there are quite a few languages in which this effect seems to have been phonologized, for example in the form of rules that require only short vowels in closed syllables or forbid them in open syllables.

Phonetic vowel duration before single and geminate consonants

The best test for a relationship between vowel length and syllabification is to be found in languages which allow word-internal geminated intervocalic consonants. It is assumed that the analysis of these geminates is as a sequence of two identical consonants with a syllable boundary falling between them. The syllable closed by the first of a pair of geminated consonants can be compared with a single consonant of the same type which is the onset to a second syllable, i.e. $C_1V_1C_2.C_2V_2$ compared with $C_1V_1.C_2V_2$. In this way the contrast is limited to only the syllabic structure and all other variables are controlled for.

A shorter vowel before geminate than before single consonants is known to occur at least in Kannada, Tamil, Telugu, Hausa, Italian, Icelandic, Norwegian, Finnish, Hungarian, Arabic, Shilha, Amharic, Galla, Dogri, Bengali, Sinhalese, and Rembarrnga. I will review some of the phonetic data from these languages, reporting principally on studies in which measurements from several speakers are provided.

Mona Lindau (personal communication) has measured the duration of short vowels before single and geminate consonants in three pairs of words in the Chadic language Hausa. The results are given in Table 1.

VC-		VCC-		V:C-	
word	V duration	word	V duration	word	V duration
cítàa	67	cíttàa	46	líitàa	106
wátàa	71	báttàa	64	fáatàa	118
gádàa	67	háddàa	50	táadàa	125
means	68		53		116
	difference	15			

Table 1. Short vowels before single and geminate plosives in Hausa. Long vowel duration is provided for comparison. Each value is the mean of 2 or 3 tokens from 9 or 10 speakers, except for the third set for which data from only six speakers is available.

Italian also has shorter vowels before geminate consonants (Antonetti and Rossi 1970). The difference in duration is greater in Italian than it is in Hausa. Some measurements of the vowel /a/ before single and geminate affricate /tS/ were made at two different speech rates by Maddieson (1980). The results are reproduced in Table 2 below.

	slow speech rate		fast speech rate	
	VC-	VCC-	VC-	VCC-
	208	132	153	112
difference	76		41	

Table 2. Vowel duration before single and geminate affricates in Italian at two speech rates. Each value is a mean of 10 tokens from each of 8 speakers.

Ghai (1980) recorded data from 5 speakers of Dogri, an Indo-Iranian language related to Panjabi. He reports that the short vowels /ə/, /ɪ/, and /ʊ/ before the geminates were 27 msec on average shorter than the vowels before single consonants in the three word pairs in Table 3, and the long vowels /o:/ and /ɑ:/ were 36 msec shorter before geminates in the two word pairs in Table 3. (The mean values themselves are not reported, only the differences.)

short vowels	long vowels
ˈkəca: / ˈkəcca:	ˈbo:li: / ˈbo:lli:
ˈkʊli: / ˈkʊlli:	ˈʄɑ:da: / ˈʄɑ:dda:
ˈkɪla: / ˈkɪlla:	

Table 3. Word pairs with single and geminate consonants in Dogri used in Ghai (1980).

It should be noted that there is a tendency for single medial stops to become fricatives in this language. The examples cited here, with the possible exception of /ʄɑ:da:/, are exempt from this trend.

In the variety of Icelandic labeled Norðlenzka (Northern Icelandic) by Oreŝnik and Pétursson (1977) vowel quantity can be predicted. For our purposes what is important is that the vowels are about half as long before long voiceless stops (orthographic bb, dd, gg) than before single voiceless stops (orthographic p, t, k). The mean duration of all vowels classed as short is 81 msec, the mean duration of long vowels is 163 msec. Data are from two speakers, but results are not given separately for vowels before geminates as opposed to other clusters or environments in which short vowels are found. (The facts in the more standard Southern Icelandic are different and call for vowel length to be taken as an underlying contrast, see Oreŝnik and Pétursson 1977: 163-167.) In another North Germanic language, Norwegian, Fintoft (1961) found that vowels before geminate consonants had a mean duration 94 msec shorter than vowels before single consonants in a set of nonsense words (8 speakers). That CSVS also operates in Finnish can be deduced from a remark by Wiik (1965): "The same amount of lengthening [as is found when comparing final open syllables with final closed syllables] is found in words like muuta, puuta as compared with muutta, puutta" (p. 118). Wiik has data from 5 speakers but does not publish these measurements separately.

Less extensive data is available on several additional languages. Length differences consistent with CSVS can be seen in spectrographic data from the Dravidian language Telugu, although this evidence is only from a single speaker (Peri Bhaskararao, personal communication). Balasubramanian (1972) indicates that shorter vowels occur before the geminate sonorants that remain in Tamil. Applegate (1958: 13) reports shortening of vowels before geminates in Shilha (Berber) on the basis of spectrographic data from one speaker. McKay (1980) found in spectrograms of one speaker of the Australian language Rembarrnga that "in general shorter vowels occurred before the geminate stops than before the single stops". Informal examination of material in the language data archives of the UCLA Phonetics Laboratory confirms that the same phenomenon is found in Amharic, Galla, Kannada (cf. Gowda 1970), Bengali, Sinhalese, and Arabic. Surprisingly no published measurements on this question in Arabic could be located, although a few spectrograms are included in Al-Ani (1970) showing shorter vowels before /ʔʔ/ and /ʕʕ/ than before /ʔ/ and /ʕ/.

Thus the reality of CSVS can be demonstrated in data from languages of several diverse language families which provide the most controlled environment for its observation, namely, before geminate and single consonants. It can be shown to occur in languages with and without a lexical vowel length contrast, in different speech rates and under different prosodic conditions.

In at least one language the phonetic length difference before single and geminate consonants is in the process of being converted into what is essentially a phonological contrast of vowel length. In the cornouallais dialect of Breton studied by Bothorel (1982) the distinction between single and geminate sonorants, preserved in the léonais dialect, has been reduced to insignificance. Measurements of both consonant and preceding vowel durations are given in Table 4. The difference in consonant duration is an insignificant 5 msec, but the vowels before the former geminates are 40 msec shorter than before the historical single consonants.

*-VC-		*-VCC-	
V duration	C duration	V duration	C duration
127	48	87	53

Table 4. Mean vowel length before former single and geminate sonorants and duration of following consonant in a Breton dialect, calculated from data in Bothorel (1982). (3 speakers, 5 words of each type).

In some Dravidian languages, such as Tamil and Malayalam, the formerly general contrast between single and geminate consonants has been eliminated from obstruents and replaced by a contrast between long voiceless stops and short voiced fricatives (Lisker 1958; Velaydhan 1971). In these languages the vowel length difference before the former contrasting single and geminate stops is retained. Tamil data from 4 speakers is given by Balasubramanian (1981). Means from his results are reproduced as Table 5.

	short vowels				long vowels			
	VC-	VCC-	CVC-	CVCC-	V:C-	V:CC-	CV:C-	CV:CC-
mean	112	97	93	80	221	188	184	152
diff.	15		13		33		32	

Table 5. Mean durations of 7 long and short vowels before single and geminate voiceless plosives in Tamil (4 speakers, about 60 tokens per speaker), after Balasubramanian (1981).

In the case of Swedish a different phonological consequence has ensued from the restructuring of length contrasts originally related to syllable structure into quantitative and qualitative distinctions between sets of "tense" and "lax" vowels. Elert (1964) provided measures across different vowel pairs before single and geminate consonants from 8 speakers. His results before /t/ and /tt/ in the two Swedish word accent patterns are reproduced in Table 6.

	Accent I		Accent II	
	-VC-	-VCC-	-VC-	-VCC-
	140	90	134	92
difference	50		42	

Table 6. Vowel duration before single and geminate alveolar stops in Swedish in two accent patterns. Each value represents a mean of 10 tokens for each of a set 9 vowels. (Data recalculated by Mona Lindau from Elert's raw measures).

Phonological constraints on vowel quantity and consonant gemination

Elsewhere, other reflections of the association between shorter vowel and geminate consonant can found in phonotactic constraints. Quite commonly those languages with both long and short vowels and single and geminate consonants restrict the vowels before geminate consonants to being phonologically short.[6] This rule is found in Arabic (Al-Ani 1970), Hausa (Abraham 1959), Hindi (Ohala 1972), Estonian (Lehiste 1966), and apparently in both Gowda and Standard dialects of Kannada (Gowda 1970) and Ulithian (Sohn and Bender 1973). In Koya "long vowels do not occur before geminates" (Tyler 1969: 6), and this language also has morphophonemic rules that shorten an underlying long vowel when geminates are derived. In Punjabi the set of "lax" centralized vowels, which tend to be shorter than the peripheral vowels, are the only vowels which may precede geminate consonants (Gill and Gleason 1963: 12). In a Bavarian dialect of German, Bannert (1972) reports that a long vowel can only precede a short consonant and a

short vowel can only precede a long consonant (in the minimal pair [vi:sn] vs. [vis:n] vowel length is 190 msec vs. 110 msec).

We thus see that both on the phonetic level and in phonological constraints shorter vowels frequently precede a geminate consonant that contrasts minimally with a single consonant within a word. In other words, the shorter vowel is in the closed syllable. An apparent counterexample, Japanese, will be discussed below. Otherwise all the languages on which data is to hand show the occurrence of a shorter vowel in a syllable closed by a geminate consonant.

Vowel duration in open and closed syllables generally

In a fairly wide range of other languages there are phonetic measurements available or brief descriptive remarks that indicate that there is a difference between vowel length in open syllables and closed syllables in general. Jones (1950: 126-128) was among the first to measure such differences in English, comparing such items as see, with seed and seat. Jones also comments on a similar difference for Russian (p. 132). Wiik (1965) confirmed Jones' findings on a larger scale for English and extended them to Finnish, using 5 speakers from each language. Han (1964: 57-61) reported that in a set of Korean data (29 words from each of 4 speakers) the mean duration of the vowel /a/ in CV syllables was 266 msec, whereas in CVC syllables it was 127 msec. In Standard Chinese the only possible syllable-final consonants are nasals. Mean values for a set of vowels and diphthongs measured by Ren Hong-Mo (personal communication) before /n/ and /ŋ/ are 238 and 200 msec respectively, whereas these syllable nuclei without a final nasal are 363 msec long (means of data from 4 speakers). Phonologically long vowels in closed syllables in Thai are reported as substantially shorter than the same vowels in open syllables (Abramson 1962). Listeners' judgments of vowel quantity reflect an awareness of this fact in that a shorter vowel is judged to be phonologically long in a closed syllable than is the case in an open one.

Brief remarks on vowel duration and syllabification in other languages include the following. For Assamese, Goswami (1966: 114) reports that stressed vowels are longer in open syllables than elsewhere in nonfinal syllables. Buth (1980) reports that "long vowels are lengthened slightly in open syllables" in the Nilotic language Jur Luo.

A phonologization of the kind of distribution of vowel duration discussed here can be found in several languages, either as a synchronic or historical process. For example, in Ngizim there is a general limitation on vowels in closed syllables, requiring them to be phonologically short (Schuh 1978: 255). In an earlier period of English, short vowels in open syllables were lengthened, in some cases merging with existing phonologically long vowels (principally the lower vowels /ɛ, a, ɔ/ were affected). The phonetic basis of this change, namely the correlation of length and syllable structure, inspired the 13th-century monk Orrm to devise an orthography in which all short vowels were indicated by writing a geminate consonant after them. (For a convenient brief summary of these facts see Strang 1970).

Hence, several further languages which show a general relationship between shorter vowel and closed syllable can be added to those which provide evidence for the widespread effects of CSVS. If CSVS is universal there will be no languages in which it does not occur. Therefore a search for possible counterexamples was conducted.

Apparent counterexamples to CSVS

There are a small number of apparent counterexamples to CSVS. One of these, Japanese, is the only documented language which shows no difference in the length of the vowels preceding geminate and single consonants in word-medial position (Han 1962, Homma 1981, Dalby and Port 1982). However, Japanese has long been held to be a language which is organized temporally on the basis of the mora (see, for example, Bloch 1950). That is /kan/ is a two mora word equivalent in this respect to /kana/, but /kana/ is not equivalent to /kanna/ which has three moras. Many analyses of Japanese treat the word-final consonants and the first element of the "geminates" as syllabic consonants (e.g Jorden 1963). The first part of a a geminate consonant in Japanese derives from a former CV syllable and is represented orthographically by a symbol which corresponds to such a syllable (Miller 1967: 109). In an emphatic (?facetious) style of pronunciation this syllable may be pronounced in full (Akira Fukuyama, personal communication). The two elements of a geminate are also separated in a Pig Latin-type secret language, reflecting a division such as /ka.n.na/. There is thus specific evidence in the case of Japanese to reject the general assumption made above that the first part of a geminate is the coda of the syllable containing the preceding vowel. Japanese is therefore not a genuine counterexample to CSVS. All other languages with geminates that have been studied show shorter vowels before them.

The other apparent counterexample to the relation between phonetic vowel length and consonant gemination arises from the conclusion of Delattre (1968) that "in distinguishing a geminate from a single consonant, the duration of the preceding vowel is a negligible factor" in English, French, Spanish and German (p. 126). These are not languages which have geminates in a similar sense to those which are found in the languages surveyed above. Most of the examples used concern consonants which occur on either side of word boundaries. Delattre comments that "what is most striking as one looks at spectrograms of these utterances is the number of cases in which a vowel preserves its original length despite a practical doubling of the following consonant's duration, as in The race ends vs. The race sends." (p.126). In these sentences [ei] is 170 msec long in each case but [s#] is 120 msec in the first while [s#s] in the second is 230 msec. But there is no reason to consider the final consonant of race in the first sentence to have been resyllabified as an onset to the word ends; in such circumstances many English speakers have a distinct word-initial onset to the vowel with glottalization. Hence there is no reason to anticipate a longer vowel in race in this sentence. Delattre's data do not address the issue of concern in this paper.

As for counterexamples to the more general correlation of shorter vowel with closed syllable, there is a possible one in French. Standard descriptions of French (e.g. Martinet 1960) mention that there is a rather limited length contrast of /ã/ vs. /ã:/ on the basis of such contrasts as grand vs. grande (/grã/ vs. /grã:d/). The long vowel occurs in the form with the closed syllable. Malmberg (1964) challenges this account, suggesting that the longer vowel optionally occurs as an indication of the originally closed nature of the syllable when it is resyllabified, as in an example such as la grande Adèle where the syllabification would be /la.grã.da.dɛl/, but that it is otherwise absent. No other authorities seem to agree with Malmberg. In fact, the long vowel in forms like grande probably originally arose when the feminine became disyllabic with addition of /-e/ (Ewart 1943) and thus had two open syllables at a time when the masculine grand was a single closed syllable. Hence the /ã/ vs. /ã:/ contrast derives from operation of CSVS, and Martinet's account of it as an underlying contrast (for those speakers who maintain it) is probably preferable.

Some languages have phonological rules that appear to run counter to CSVS. One language which is reported to have a phonological rule that lengthens vowels in closed syllables is the Micronesian language Kusaiean (Lee 1975, Levin 1983). In one pattern of reduplication a short vowel in the unreduplicated form is repeated as a long vowel in a closed syllable, e.g. the simple form /fule:/ has the derived form /fu:lfule:/. The reduplication process appears sensitive to the closed nature of the syllable because there is another reduplication pattern in which the medial consonant is not repeated, that is, the reduplicated syllable is CV rather than CVC. In this pattern when the simple form has a short vowel the reduplicated syllable also has a short vowel, giving for example /fufule:/. However, it should be noted that the lexically short vowels in this language are severely restricted in their distribution and it is clear that vowels are "normally" long. The short vowels are only permitted in the first syllable of disyllabic or longer words (apart from a few derived forms including some reduplicates). All vowels in monosyllables and noninitial syllables are long. Moreover only nonlow vowels may be short. Lee's grammar of Kusaiean does not include any evidence to suggest that the occurrence of the short vowels can be predicted, but their limited distribution does suggest that this might be a possibility. An account of Kusaiean in which all vowels are underlyingly long would replace the rule that lengthens vowels in closed reduplicated syllables with one that shortens vowels under certain conditions that are not tied to syllable structure. It is unclear that such an account can be successfully made, but if it were it would have the advantage of explicitly representing the fact that the long vowels are the normal variants.

According to Bloomfield (1939), Menomini has rules which both lengthen short vowels in closed syllables and shorten long vowels in open syllables under certain conditions. These rules are a part of a process which appears to be mainly concerned with establishing an alternating rythmic pattern, which is in part tied in with stress (Pesetsky 1979). The rythmic pattern seems to evaluate CV and CVVC syllables as equivalents; the even-numbered syllables following the first long vowel in a word are changed minimally so that they conform to one or the other of these structures. On the other hand, the vowel in the second syllable of a word is lengthened regardless of whether it is in an open or closed syllable, and vowels in the odd-numbered syllables are unchanged in length. This set of rules taken as a whole does not produce any general association of length and syllable structure which is counter to CSVS.

The above are the possible counterexamples to CSVS that I am aware of. They do not seem to be such as to seriously challenge the validity of the claim that CSVS is found across the broad generality of languages.

Discussion

CSVS seems to be present in the world's languages with sufficient uniformity that it can be used as a cue to the syllabic constituency of a string of segments. In addition, the demonstration of the generality of of CSVS may have an important implication for understanding of speech production and linguistic structure. CSVS is consistent with the view that the rhyme of a syllable is a unit of organization in speech production. This view is connected with but not identical to the view that -VC- sequences are units of timing organization. Many studies have drawn attention to the tendency for vowel duration to be longer and consonant duration to be shorter in VC sequences in which the following consonant is voiced compared to when it is voiceless. Walsh and Parker (1982) have used this inverse relationship as evidence for the unity of the VC portion of a CVC

syllable (at least at some point in the derivation). Port (1981) also argues for a similar "macrounit" consisting of the "vowel plus following tautosyllabic consonant" (p. 272). However much of the experimental data on vowel duration and consonant voicing demonstrates that voicing-related length variations occur whether or not there is an intervening syllable boundary. Chen (1970) provides examples from French and Russian which show these phenomena before both tautosyllabic and heterosyllabic consonants, and Korean data in which all the consonants are heterosyllabic but which still shows a mean vowel duration 28 msec shorter before voiceless (aspirated) stops than before voiced stops. In Balasubramanian's Tamil data mean duration of long vowels before heterosyllabic voiced consonants is 30.3 msec longer than before voiceless ones; short vowels are 14 msec longer before voiced consonants than before voiceless ones (data recalculated from tables in Balasubramanian 1981).

None of the many studies on this question of vowel length before consonants which contrast in voicing report data in a way that enables the possible effect of syllable boundary placement to be entirely separated from other factors which also affect duration (such as word length, consonant manner, vowel quality, etc.). However, in Chen's Russian data the 5 word pairs in which the vowel was measured before a heterosyllabic consonant have a mean difference of 27.2 msec, whereas the 6 word pairs with tautosyllabic consonants show a mean difference of 29.8 msec. Consonants and vowels are not matched across these two sets of words and the majority of tautosyllabic cases occur in monosyllables whereas the heterosyllabic cases are in disyllables. Furthermore, there is a rule of final obstruent devoicing in Russian; hence the monosyllables examined by Chen show a contrast in which the phonetic presence of voicing is probably not a factor.[7] Nonetheless there is some indication here that the length difference associated with the voicing contrast is the same in open syllables and in closed syllables.

Because of this, those who have argued that the vowel (and consonant) length adjustment associated with voicing contrast provides evidence for the unity of VC as a constituent of the syllable have yet to show that there is any basis for doing so. This influence on vowel length behaves like certain types of coarticulation, such as anticipatory rounding of the lips, which has been shown to ignore word (Bell-Berti and Harris 1982) and syllable boundaries (to judge from McAllister 1978). Although data from coarticulation studies has sometimes been interpreted as throwing light on the syllabic organization of speech production (e.g. Song and Perkell 1983), these studies do not normally examine utterances which differ minimally in syllabification. Hence they also do not address the question of syllable structure in speech production and language representation.

CSVS on the other hand is (ex hypothesi) related to syllable structure and thus provides a basis for drawing conclusions about the role of the syllabic unit in languages and in the general human capacity for producing articulate speech. It also provides some support for those such as Kiparsky (1979) and Selkirk (1980) who wish to recognize the rhyme as an internal constituent of the syllable.

Notes

1.
This is implicit, for example, in Clements and Keyser's (1983) discussion of their Resyllabification Convention (RC) which states that "the output of every

rule is resyllabified according to the syllable structure rules examined up to that point in the derivation" (p. 54). They go on to comment that "individual grammars may specify a point in the set of ordered rules at which the [RC] becomes inoperative; indeed, some languages may not make use of the [RC] at all. .....Resyllabification across word boundaries.....is normally optional, and may differ in some respects from initial syllabification."

2.
Jones (1931 [1972]) asked a similar question about the phonetic reality of the word and drew attention to several features which serve to demarcate words in English, including marking the internal boundaries in compounds.

3.
Even British English accents differ in this regard. For example, in Jenny Ladefoged's speech "holy" and "holey" rhyme, both being [hou.li]. Note that in my speech the lateral in "holey" could well be considered to be an ambisyllabic consonant. However, this means that it is still a constituent of the first syllable, which is the major point of interest here, so I will not take space to argue for or against this interpretation.

4.
Note that the example in figure 2 shows /tsa:/ before underlying /b/ and /p/. If this underlying voicing difference made a contribution to the difference in vowel length in these two phrases it would be expected to be in the opposite direction from that seen, i.e. vowels are commonly longer before voiced stops (for a brief discussion of the generality of this phenomenon see Javkin 1979: 53). Ohala (1981) suggests use of this length distribution as a tool to determine if [p] in words like teamster [timpstɚ] is underlying or intrusive. Such a use of phonetic patterning is similar to using the vowel length difference as evidence for syllabification, as suggested here and in Maddieson (1983).

5.
Another property of vowels which is associated with the difference between closed and open syllables and which may possibly be linked to the question of length is the tendency for high and mid vowels to have lower and/or more central allophones in closed syllables than in open syllables. This may also help to determine whether a postvocalic consonant is closing the syllable in which the vowel occurs or is the onset to a following syllable. At this point the generality of this tendency is harder to establish than the generality of CSVS. Some effects of vowel lowering in closed syllables are reported in at least Kharia (Pinnow 1959), Kurukh (Pinnow 1964), Javanese (Herrfurth 1964), Danish (Basbøll 1974), Spanish (Navarro Tomás 1968) and French (Lennig 1978). Like CSVS, this tendency has also left its historical imprint in a number of languages. A well-known example is French (Ewart 1943: 42-48), where phonetic changes based on the open or closed nature of the syllable have occurred in several historical periods. One result of this is the kind of alternation seen in forms like sot, sotte [so, sɔt] and espère, espérons [ɛspɛr, ɛsperɔ̃]. CSVS and vowel lowering in closed syllables could be linked because of the general association between shortness and lower or more central quality in vowels which is found in languages as diverse as Navaho, Kurdish, Arabic and Somali. For some discussion of the relationship see Straka (1959).

6.
No restriction of this kind operates in Tamil and Finnish where we have already seen that the phonetic shortening effect is found, nor in Hungarian, Malayalam

and various other languages where phonetic data is not readily to hand. (Hegedüs (1958) published waveforms of words with geminates in Hungarian but provides no data on contrasting words with single consonants.)

7.
It is not clear if phonologically voiced obstruents which have been devoiced are actually phonetically the same as the phonologically voiceless obstruents in this position in Russian. It seems likely that they are not, just as the stops in pairs of English words like "rope" and "robe" are not.

References

Abercrombie, D. 1967. Elements of General Phonetics. Edinburgh University Press, Edinburgh.

Abraham, R.C. 1959. The Language of the Hausa People. University of London Press, London.

Abramson, A.S. 1962. The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments. Indiana University, Bloomington.

Al-Ani, S.H. 1970. Arabic Phonology. Mouton, The Hague.

Applegate, J. R. 1958. An Outline of the Structure of Shilha. American Council of Learned Societies, New York.

Antonetti, P. and Rossi, M. 1970. Précis de Phonétique Italienne: Synchronie et Diachronie. La Pensée Universitaire, Aix-en-Provence.

Balasubramanian, T. 1972. The Phonetics of Colloquial Tamil. Ph.D. Dissertation, University of Edinburgh.

Balasubramanian, T. 1981. Duration of vowels in Tamil. Journal of Phonetics 9: 151-161.

Bannert, R. 1972. Zur Stimmbehaftigkeit und Quantität in einem bairischen Dialect (Working Papers 6). Phonetics Laboratory, University of Lund.

Basbøll, H. 1974. The phonological syllable with special reference to Danish. Annual Report of the Institute of Phonetics, University of Copenhagen 8: 39-128.

Bell-Berti, F. and Harris, K.S. 1982. Temporal patterns of coarticulation: lip rounding. Journal of the Acoustical Society of America 71: 449-454.

Bloch, B. 1950. Studies in colloquial Japanese: IV phonemics. Language 26: 86-125.

Bloomfield, L. Menomini morphophonemics. Travaux du Cercle Linguistique de Prague 8: 105-115.

Bothorel, A. 1982. Etude phonétique et phonologique du Breton parlé à Argol (Finistere-sud). Atelier National Reproduction des Thèses, Université Lille III, Lille.

Buth, R. 1981. The twenty vowels of Dhe Luwo (Jur Luwo, Sudan). Proceedings of the First Nilo-Saharan Linguistics Colloquium, Leiden (ed. by T. Schadeberg and M.L. Bender). Foris, Dordrecht: 119-132.

Cairns, C.E. and Feinstein, M.H. 1982. Markedness and the theory of syllable structure. Linguistic Inquiry 13: 193-226.

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

Clements, G.N. and Keyser, S.J. 1983. CV Phonology: A Generative Theory of the Syllable. MIT Press, Cambridge, Mass.

Dalby, J. and Port, R.F. 1981. Temporal structure of Japanese: segment, mora and word. Research in Phonetics (Dept. of Linguistics, Indiana University, Bloomington) 2: 149-172.

Delattre, P. 1968. Consonant gemination in four languages: an acoustic, perceptual and radiographic study. The General Phonetic Characteristics of Languages: Final Report 1968 (University of California, Santa Barbara): 105-163.
Elert, C-C. 1964. Phonologic Studies of Quantity in Swedish. Skriptor, Uppsala.
Ewart, A. 1943. The French Language (2nd. edition). Faber, London.
Faure, G. 1972. Analyse acoustique de deux allophones du l final anglais. In Papers in Linguistics and Phonetics to the Memory of Pierre Delattre (ed. by A. Valdman). Mouton, The Hague: 117-127.
Fintoft, K. 1961. The duration of some Norwegian speech sounds. Phonetica 7: 19-39.
Ghai, V.K. 1980. Contributions to Dogri phonetics and phonology. Annual Report of the Institute of Phonetics, University of Copenhagen 14: 31-94.
Gill, H.S. and Gleason, H.A. 1973. A Reference Grammar of Punjabi. Hartford Seminary Foundation, Hartford, Connecticut.
Goswami, G.C. 1966. An Introduction to Assamese Phonology. Deccan College, Poona.
Gowda, K.K. 1970. Gowda Kannada. Annamalai University, Annamalainagar.
Han, M.S. 1962. The feature of duration in Japanese. The Study of Sounds (Phonetic Society of Japan) 10: 65-75.
Han, M.S. 1964. Studies in the Phonology of Asian Languages 2: Duration of Korean Vowels. University of Southern California, Los Angeles.
Harris, J.W. 1983. Syllable Structure and Stress in Spanish: A Nonlinear Analysis. MIT Press, Cambridge, Mass.
Hegedüs, L. 1959. Beitrag zur Frage der Geminanten. Zeitschrift für Phonetik und Allgemeine Sprachwissenschaft 12: 68-106.
Herrfurth, H. 1964. Lehrbuch der Modernen Djawanischen. VEB, Leipzig.
Homma, Y. 1981. Durational relationships between Japanese stops and vowels. Journal of Phonetics 9: 273-281.
Javkin, H.R. 1979. Phonetic Universals and Phonological Change. Report of the Phonology Laboratory (University of California, Berkeley), No. 4.
Jones, D. 1931 [1973]. The "word" as a phonetic entity. In Phonetics in Linguistics: A Book of Readings (ed. W.E. Jones and J. Laver). Longmans, London: 154-158. [Transcribed from Le Maître Phonétique, 3rd. series, 36: 60-65 (1931)].
Jones, D. 1950. The Phoneme: Its Nature and Use. Heffer, Cambridge.
Jorden, E.H. 1963. Beginning Japanese: Part I. Yale University Press, New Haven.
Kahn, D. 1976. Syllable-based Generalizations in English Phonology. Ph. D. Dissertation, MIT. Distributed by Indiana University Linguistics Club, Indiana University, Bloomington.
Kiparsky, P. 1979. Metrical structure assignment is cyclic. Linguistic Inquiry 10: 421-442.
Ladefoged, P.L. 1982. A Course in Phonetics (2nd ed.). Harcourt Brace Jovanovich, New York.
Lee, K-D. 1975. Kusaiean Reference Grammar. University Press of Hawaii, Honolulu.
Lehiste, I. 1964. Acoustical Characteristics of Selected English Consonants. Indiana University, Bloomington.
Lehiste, I. 1966. Consonant Quantity and Phonological Units in Estonian. Indiana University, Bloomington.
Lennig, M. 1978. Acoustic Measurement of Linguistic Change: The Modern Parisian Vowel System. Ph. D. Dissertation, University of Pennsylvania. Distributed by U.S. Regional Survey, Philadelphia.
Levin, J. 1983. Reduplication and prosodic structure. Ms. MIT, Cambridge, Mass. (Revised version of a paper presented at GLOW Colloquium, York, March 1983. Abstract in GLOW Newsletter 10: 52-54.)

Lisker, L. 1958. The Tamil occlusives: short vs. long or voiced vs. voiceless? Indian Linguistics, Turner Jubilee Volume 1: 294-301.
Maddieson, I. 1980. Palato-alveolar affricates in several languages. UCLA Working Papers in Phonetics 51: 120-126.
Maddieson, I. 1983. The analysis of complex phonetic elements in Bura and the syllable. Studies in African Linguistics 14.
Malmberg, B. 1955 [1967]. The phonetic basis for syllable division. In Readings in Acoustic Phonetics (ed. by I. Lehiste), MIT Press, Cambridge, Mass: 293-300 [Reprinted from Studia Linguistica 9: 80-87 (1955).]
Malmberg, B. 1964. Juncture and syllable division. In In Honour of Daniel Jones (ed. by D. Abercrombie et al.). Longmans, London: 116-119.
Marlett, S.A. and Stemberger, J.P. 1983. Empty consonants in Seri. Linguistic Inquiry 14: 617-639.
Martinet, A. 1960. Eléments de linguistique générale. Colin, Paris.
McAllister, R. 1978. Temporal asymmetry in labial coarticulation. Papers from the Institute of Linguistics, University of Stockholm 35.
McKay, G.R. 1980. Medial gemination in Rembarrnga: a spectrographic study. Journal of Phonetics 8: 343-352.
Miller, R.A. 1967. The Japanese Language. University of Chicago Press, Chicago.
Navarro Tomás, T. 1968. Manuel de Pronunciación Española (14th ed.). Consejo Superior de Investigaciones Cientificas, Madrid.
Ohala, J.J. 1981. Speech timing as a tool in phonology. Phonetica 38: 204-212.
Ohala, M. 1972. Topics in Hindi-Urdu Phonology. Ph. D. Dissertation, University of California, Los Angeles.
Orešnik, J. and Pétursson, M. 1977. Quantity in Modern Icelandic. Arkiv för Nordisk Filologi 92: 155-171.
Pesetsky, D. 1979. Memomini quantity. MIT Working Papers in Linguistics 1: 115-139.
Pike, K.L. 1943. Phonetics. University of Michigan Press, Ann Arbor.
Pinnow, H-J. 1959. Versuch einer Historischen Lautlehre der Kharia-Sprache. Harrassowitz, Wiesbaden.
Pinnow, H-J. 1964. Bemerkungen zur phonetik und phonemik des Kurukh. Indo-Iranian Journal 8: 32-55.
Port, R.F. 1981. Linguistic timing factors in combination. Journal of the Acoustical Society of America 69: 262-274.
Schuh, R.G. 1978. Bade/Ngizim vowels and syllable structure. Studies in African Linguistics 9: 247-283.
Selkirk, E. 1980. The role of prosodic categories in English word stress. Linguistic Inquiry 11: 563-605.
Sohn, H-M. and Bender, B.W. 1973. A Ulithian Grammar (Pacific Linguistics, Series C, 27). Australian National University, Canberra.
Song, S.S. and Perkell, J.S. 1983. A syllabic component of speech motor control? Speech Communication Group Working Papers (Research Laboratory of Electronics, MIT) 2: 67-76.
Steriade, D. 1982. Greek Prosodies and the Nature of Syllabification. Ph.D. Dissertation, MIT, Cambridge, Mass.
Stetson, R.H. 1951. Motor Phonetics (2nd ed.). North Holland, Amsterdam.
Straka, G. 1959. Durée et timbre vocalique. Zeitschrift für Phonetik und Allgemeine Sprachwissenschaft 12: 276-300.
Strang, B.M.H. 1970. A History of English. Methuen, London.
Tyler, S.A. 1969. Koya: An Outline Grammar. University of California Press, Berkeley and Los Angeles.
Velayudhan, S. 1971. Vowel Duration in Malayalam: An Acoustic Phonetic Study. Dravidian Linguistic Association of India, Trivandrum.

Walsh, T. and Parker, F. 1982. Consonant cluster abbreviation: an abstract analysis. Journal of Phonetics 10: 423-438.
Wiik, K. 1965. Finnish and English Vowels. University of Turku, Turku.

Using a Spectrograph for Measures of Phonation Types in a Natural Language

Paul L. Kirk (Department of Anthropology, California State University, Northridge), Peter Ladefoged and Jenny Ladefoged (Phonetics Laboratory, Linguistics Department, UCLA)

Obviously the best way of evaluating what the vocal cords are doing is to look at them. If we want to quantify different types of phonation then the most direct technique is to measure the glottal movements observed by means of high speed cinematography, and the muscular actions as recorded by electromyography. But speech scientists frequently have at their disposal only an ordinary tape recording, often one made in the field rather than under ideal laboratory conditions. In these circumstances they must quantify different types of phonation simply by reference to acoustic data. Furthermore, for many speech scientists, the only instrument available is a sound spectrograph. It is therefore important to evaluate the ways in which this instrument can be used for measuring different types of phonation recorded in field conditions.

The Kay Digital sound spectrograph is a delightful new aid for researchers in speech science, providing a number of different types of display. It will produce the traditional three dimensional spectrogram, showing the relation between amplitude, frequency, and time, using filters of various band widths. There are also facilities for obtaining the power spectrum at two precisely located points in an utterance. In addition the waveform may be displayed for the whole utterance, or more importantly, for an expanded part of the utterance. The principal aim of this research is to evaluate the effectiveness of these different displays in quantifying differences in voice quality.

The term voice quality is used in a variety of ways. Laver (1980) uses it to describe any long term characteristic of a speaker, including things such as persistent pharyngealization or labialization. We will use the term voice quality in a more restricted way, applying it only to those aspects of speech that are due to the action of the vocal cords. Differences of phonation type of this kind are sometimes referred to as breathy voice, creaky voice, laryngealized voice, etc. The major problem with these labels is that they are used in different ways by different people. Furthermore, they are often used to identify purely individual characteristics of voices. Obviously it would be preferable to have a set of terms that could be used to describe not just one voice at a time, but a group of voices each of which could be said to be, for example, breathy, meaning that each of these voices was perceived by <u>all</u> listeners to have something in common with all the other voices in the group, and this quality sets them apart from voices that might be said to have modal voice, i.e. normal voiced sounds of the type present in all languages. It would be particularly appropriate if we could find something that enabled not just trained listeners, but any listeners to say that a given pair of sounds differed in a particular voice quality.

These requirements are satisfied by using material from languages which distinguish meanings by changes in voice quality. Most languages distinguish between just voiced and voiceless sounds, and it does not matter whether the voice sounds breathy or harsh. If one speaks with very loose vibrations of the vocal cords, allowing a greater airflow than normal, a laryngologist might prescribe therapy; but such a voice quality does not affect the linguistic contrasts between words. There are, however, a number of languages in which differences in the mode of vibration of the vocal cords reflect differences in

meaning (Ladefoged 1983). When a speaker of such a language uses a change in voice quality to produce a change in meaning, then everyone who understands the language has to be able to recognize this change. Moreover all speakers have to make the change in much the same way. These languages therefore offer material that can be used for evaluating methods of measuring phonation types because one knows, a priori, that a given pair of sounds will differ in a given way. Individual speakers may have their own idiosyncratic characteristics; but for each speaker, if the meaning requires a word with a breathy voiced vowel in one instance and a modal voiced vowel in another, then the measurable difference between the two words is a measurable difference between breathy voice and modal voice. It is still, of course, possible that what we call breathy voice in one language may not be the same as what we call breathy voice in another. However, if we find that the measures we use for quantifying the difference in one language are also reliable indicators of the difference in other languages, then we can conclude that we are in fact measuring difference between phonation types that can be defined.

There is, however, a cost to using real language material that does not apply to artificially produced samples of different phonation types. Thus many clinical approaches to the study of voice quality (e.g. Davis 1976) use samples elicited by asking the patient to produce a steady state vowel lasting for several seconds. In real languages the differences occur in running speech; consequently they may be difficult to analyze because they may last for only a small part of a second. But this cost is more than offset by the advantage of being able to use several speakers, knowing that they must all be behaving in some similar ways, if they are producing words which their listeners can identify correctly.

Language data

The language we chose to examine is the Jalapa de Diaz dialect of Mazatec (hereafter referred to as Jalapa Mazatec), spoken by approximately 8,000 people in the District of Tuxtepec, Oaxaca, Mexico. Twenty three dialects of Mazatec have been identified (Kirk 1970) and a number of similarities as well as differences among the dialects are clearly observable (Kirk 1966). Mazatec is a member of the Popolocan branch (Gudschinsky 1959) of Otomanguean (Rensch 1976). The syllable structure of Huautla de Jiménez Mazatec, described by Pike and Pike (1947), has many similarities to other dialects of Mazatec (Pike 1956, Kirk 1966, Jamieson 1977a, Jamieson 1977b, Schram and Pike 1978).

Among the various dialects of Mazatec we chose Jalapa Mazatec because it is rich in voice quality distinctions. In fact it probably makes a greater use of differences in phonation types than any other language that we know of. In consonants, it makes use of voiced and voiceless oppositions as well as aspirated and unaspirated oppositions. In addition Jalapa Mazatec makes use of modal voice, creaky voice, and breathy voice for lexical differentiation. It distinguishes words such as "horse" and "arse" by the former having breathy voice and the latter having creaky (all other features of these words are the same, i.e., [ndæ̤^{1}] "horse", [ndæ$_{1}$] "arse"). Jalapa Mazatec is a tone language contrasting three pitch levels and glides between them. Contrastive tones differentiate meanings of words as well as grammatical categories and along with nasalization are the domain of the syllabic nucleus (Pike and Pike 1947). In this paper we write low tone with the widely used convention of superscript one. However, most publications on Otomanguean languages use the superscript numbers in reverse fashion with one representing high tone and increasing larger numbers representing lower tones.

In addition to the distinction between breathy voice and creaky voice, there is an opposition between modal voice and breathy voice as in:

[ni^{2}mæ^{2}] "bumble bee"	[mæ^{2}] "he wants"
[nda^{2}] "good"	[nda^{12}] "hard"
[ni^{2}ʔja^{2}] "house"	[ja^{2}] "he wears"
[jæ^{1}] "snake"	[tʃu^{1}jæ^{1}] "turtle"
[ni^{2}] "red"	[ja^{3}ni^{2}] "he carries on his back"
[ju^{2}] "willing"	[ti^{3}ju^{2}] "it is stopped"

There is also an opposition between modal voice and creaky voice as in:

[ʃĩ^{1}] "hunt"	[ʃĩ^{1}] "male"
[nu^{3}] "year"	[nu^{3}] "vine"
[kã^{2}] "twenty"	[kã^{2}] "alone"
[tʃa^{3}] "old"	[tʃa^{3}] "load"
[t^{h}æ^{2}] "itch"	[t^{h}æ^{2}] "sorcery"
[hi^{3}tsi^{1}] "yours"	[tsi^{1}] "nausea"

Three way opposition, though not minimal, between modal voice, creaky voice, and breathy voice is attested by the following examples:

[ja^{3}] "tree"	[ni^{2}ʔja^{2}] "house"	[ja^{2}] "he carries"	[ja^{2}] "he wears"
[nt^{h}æ^{1}] "seed"		[ndæ^{1}] "arse"	[ndæ^{1}] "horse"

Although a detailed examination of the phonology of Jalapa Mazatec is not the focus of this paper, nevertheless from the discussion thus far, it is evident that the phonology is fairly complex. Examples cited above show that differences in phonation type are not just an incidental phenomenon associated with a few words, but a pervasive factor of the language. If phonation type is taken to be a property of a syllable, there is a three way contrast between modal voice, creaky voice, and breathy voice. If it is regarded as a property of a segment, then we may say that there is an opposition between breathy voiced and modal voiced consonants, and between creaky and modal voiced vowels.

The material we used in our analysis comes from a recording of a longer word list illustrating a wide range of phonological properties. There were five speakers all of whom spoke the same dialect; in December 1982 four were recorded in Jalapa de Diaz and the fifth was recorded in Mexico City. For the purposes of the present analysis we concentrated on three words which were carefully matched for pitch and vowel quality. Two of the words [ndæ^{1}] "arse" and [ndæ^{1}] "horse" are differentiated only by creaky voice as opposed to breathy voice. The third word used in this study [nt^{h}æ^{1}] "seed" has a vowel with modal voice of the same quality and tone as the vowels with creaky and breathy voice.

Wide band spectrograms

We produced a number of different displays of this data. Figure 1 shows frequency-amplitude-time displays, using a 300 Hz bandwidth filter (wide band spectrograms) illustrating vowels with creaky voice, modal voice, and breathy voice for all five speakers. These vowels have been segmented from the words [nd$\ddot{æ}^1$] "arse", (creaky voice), [nt^{h}æ^1] "seed", (modal voice), and [nd$\ddot{æ}^1$] "horse" (breathy voice). Wide band spectrograms provide good displays of what Fant (1973) calls the F-pattern, the overall pattern of the formants during a sound. The formants are particularly clear during the creaky voice vowels in the top row, and fairly evident during the modal voice in the second row. In the breathy vowels in the bottom row the first formant is less well defined. In general the formants are in similar places in all three phonation types. There may be a tendency for the first formant to have a sightly higher frequency during creaky vowels -- a difference that would be associated with the raising of the larynx and the consequent shortening of the vowel tract during creaky phonation. We were, however, unable to use these displays to provide reliable measures of formant frequencies that could be used to quantify phonation types.

Wide band spectrograms also provide a very suitable basis for duration measurements. The displays of the vowels in Figure 1 were carefully segmented from the spectrograms of the complete utterances. They begin with the first pulse of the vocal cords after the release of the consonant. It is arguable that this may not be the beginning of the vowel, as the spectrograms show the formants associated with the movements of the tongue away from the alveolar ridge. These are particularly noticeable in the case of the creaky vowels in the first row. The modal vowels follow an aspirated [t^h], and the transitions are over before voicing starts. The breathy vowels in the bottom row do not have such a clear first formant, but nevertheless do exhibit some evidence of consonant transitions. Despite the presence of transitions we took the first pulse of the vocal cords as the start of the vowel because it enables us to make more reliable measures; we could not find a satisfactory procedure for defining the end of the transition.

Defining the end of the vowel was also difficult, as there is no following consonant. For the purposes of this study it was taken to be the last point at which there is energy in at least two of the formants. Given these definitions it is quite clear that both the creaky and breathy vowels are longer than the modal vowels. The mean durations for the five speakers are: creaky vowels 266 msec (s.d. 58); modal vowels 174 msec (s.d. 9); breathy vowels 244 msec (s.d. 49). Some of this difference may be associated with the differences in the contexts: the modal vowels followed an aspirated consonant cluster [nt^h], whereas the creaky and breathy vowels followed a voiced cluster [nd]. But our observations of the length differences in the other contrasts reported above also provide support for the conclusion that the breathy and creaky phonation types are regularly accompanied by greater duration. For example, the mean duration of the breathy vowel in [m$\ddot{æ}^2$] "he wants" is 317 (s.d. 40) msec, compared with the modal vowels in [ni^{2}mæ^2] "bumble bee" is 177 (s.d. 27) msec.

A number of points that are more easily quantified with the aid of other types of displays are clearly visible in the general picture provided by the wide band spectrograms in Figure 1. Most importantly it may be seen that the vowels in the first row with creaky voice are marked by jitter. The vertical striations (i.e. glottal pulses) occur at irregularly spaced intervals. For all speakers, the pattern is for pulses to be grouped closer together at the onset of the vowel

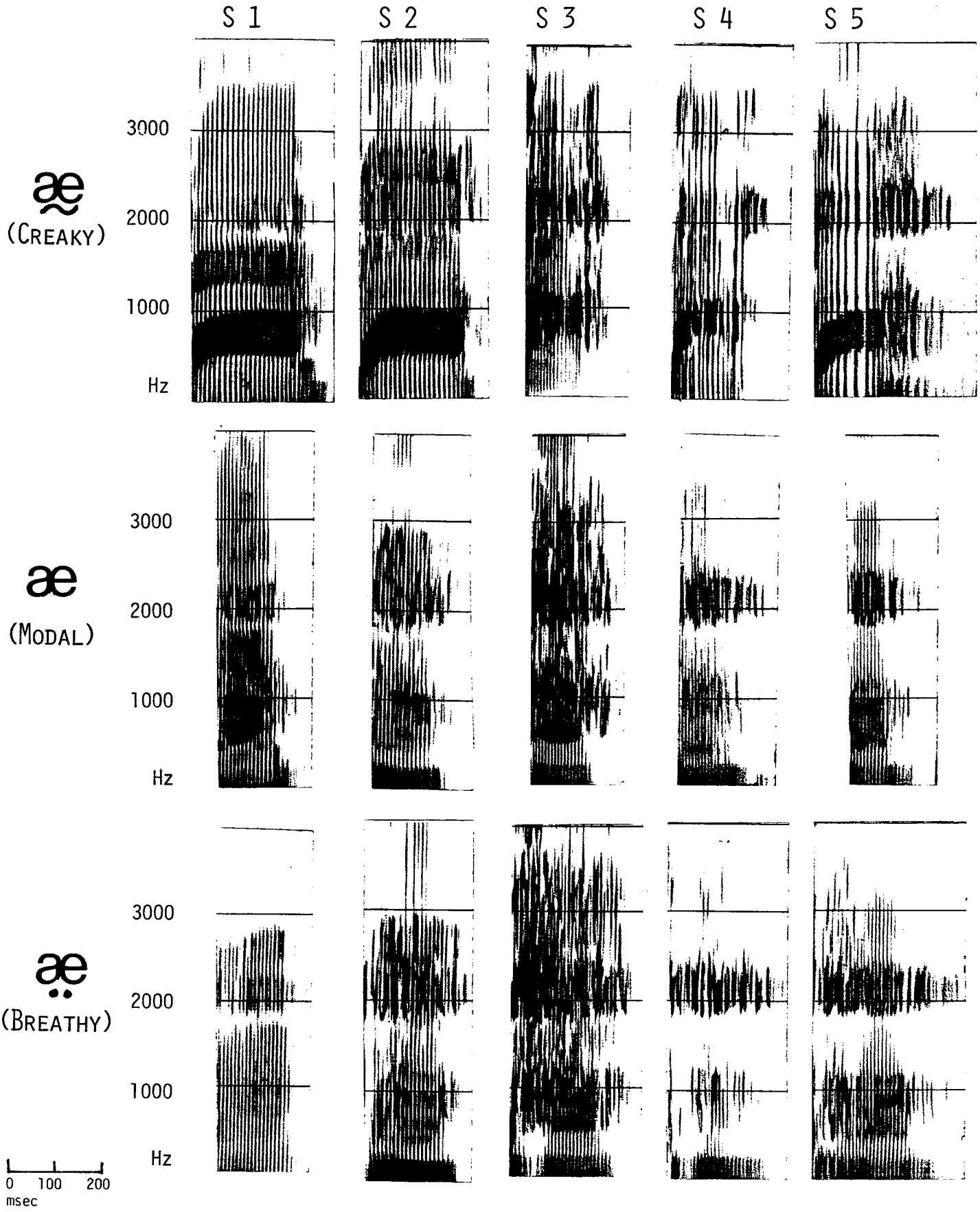


Figure 1: Wide band spectrograms of Jalapa Mazatec creaky, modal, and breathy vowels, for five speakers, S1-S5.

followed by increased distances between pulses moving toward the center of the vowel, followed by decreased distances between pulses toward the coda of the vowel.

Despite the slower rate of the glottal pulses during part of this vowel, this word is perceived by Mazatec speakers as having the same lexical tone as the modal vowel word. Mazatec speakers whistle these words with the same pitch in whistle speech. Surprisingly, phoneticians also perceive the pitch of these creaky and modal vowels as being the same. The laryngealization superimposed in the middle of the word does not offset the auditory impression created by the more regular voicing at the beginning and end of these vowels.

Qualitative differences in spectral balance are evident in Figure 1. The higher frequencies tend to be more clearly visible during the creaky vowels (see especially speakers two, four, and five). Yet another noticeable difference between the vowels in the second and third row compared with the vowels in the first row is that the bandwidth of each formant is somewhat less in the vowels with creaky voice (row one). It is also clear that the overall differences between the vowels in breathy voice (row three) in comparison with those in modal voice (row two) are similar, but in the opposite direction, to the differences that occur between creaky voice and modal voice. Furthermore, we may note that for all speakers, breathy voice is more clearly seen in the onset part of the vowel since the coda section of the vowel tends to have modal voice. Finally, we can see that for two of the five speakers, the amplitude of the first formant is distinctly less in the breathy vowels. Thus the wideband spectrograms provide an excellent general view of many aspects of differences in phonation types, some of which can be quantified by reference to other kinds of display.

Power spectra

Another kind of display that can be produced with the aid of a sound spectrograph is a power spectrum showing the relative intensities of the component frequencies. Figure 2 is a display of power spectra of the three phonation types. For these spectra, a 45 Hz (narrow band) filter was used so that the amplitudes of each of the component harmonics are clearly visible.

We noted in the discussion of the wide band spectrograms that the higher formants are more evident during creaky voice. Narrow band power spectra offer a way of quantifying this spectral tilt. But before we consider a method of measuring the relative spectral balance associated with each phonation type, we must note that there are two different ways in which higher frequencies can come to have more energy. When producing breathy voice the vocal cords are vibrating more loosely, often not making complete contact along their whole length at any time in the glottal cycle. As a result there is a greater rate of airflow through the glottis producing a turbulent airflow with more random high frequency components. When producing creaky voice the vocal cords are much more tensed, closing rapidly during each glottal cycle. As a result the vocal tract is excited by a sharper pulse that has more energy in the higher harmonics. In order to separate out these different phonation types we need a measure that distinguishes whether the increase in the higher frequencies is associated with additional components produced by a semi-random, turbulent, airflow, or by an increase in the intensity of the higher harmonics produced by a sharper glottal pulse. None of the displays produced by the spectrograph provides a way of separating out the turbulent airflow so that it can be quantified. But the power spectra enable us to quantify the relative amount of energy in different

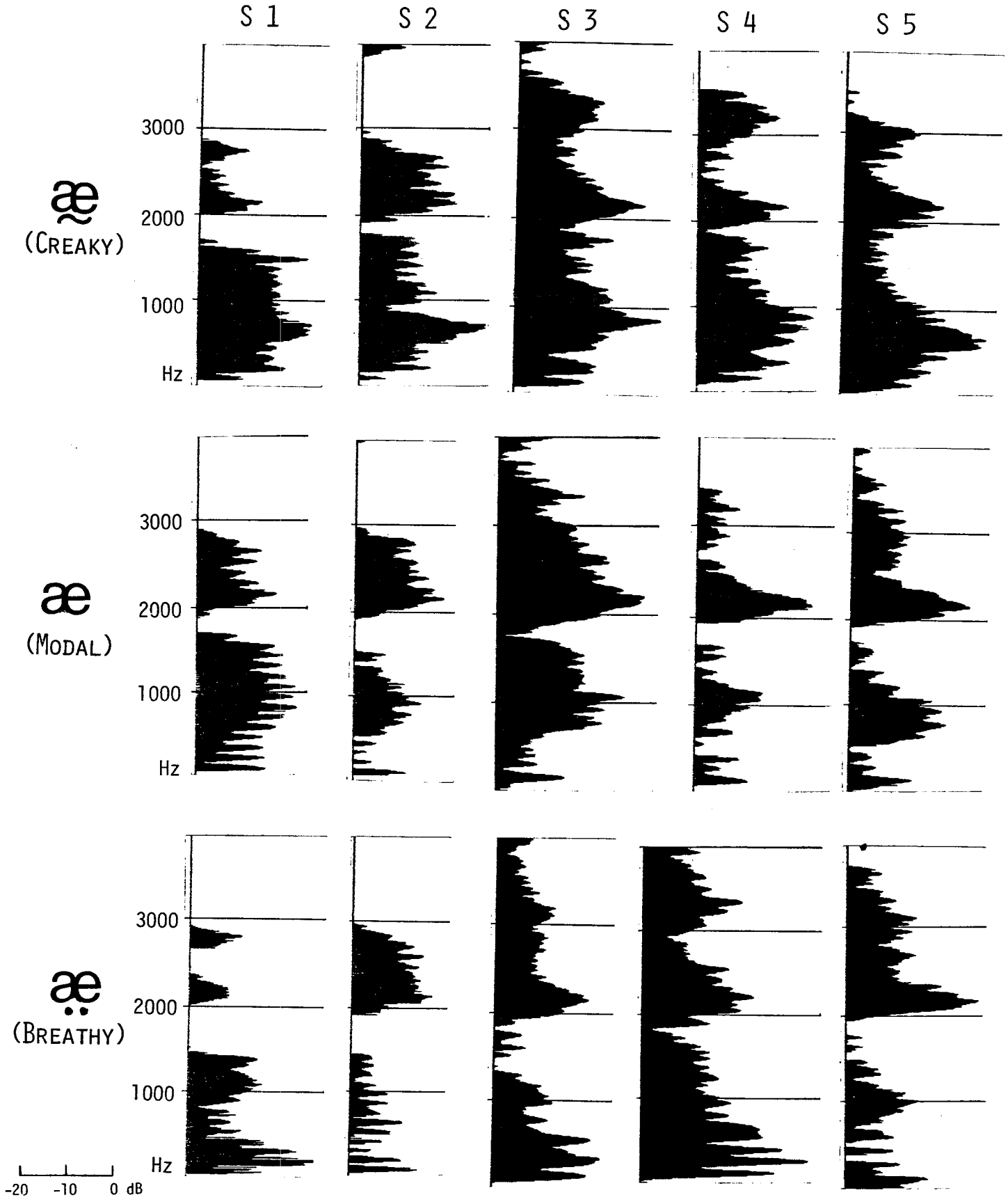


Figure 2: Spectra of Jalapa Mazatec creaky, modal, and breathy vowels.

harmonics. We chose as our measure the difference in dB between the intensity of the fundamental and the intensity of the largest harmonic in the first formant. (We are grateful to Professor K.N. Stevens for suggesting this measure to us.) This measure can be used for comparing phonation types only in cases in which the vowels being compared have similar formant frequencies; as the relative intensity of each formant is a function of its frequency (Fant 1956). But in the case of these Jalapa Mazatec vowels this is perfectly appropriate.

The difference between the amplitude of the fundamental and that of first formant is displayed in Figure 3. For the five speakers the mean for creaky voice (indicated in the bar graph with diagonal striations) is -17 dB with a standard deviation of 3.7 (i.e., the fundamental has 17 dB less amplitude than the first formant). The mean for modal voice (indicated by the black bar) is -6.6 dB with a standard deviation of 4.4. The mean for breathy voice (indicated by a cork-screw pattern) is +5.2 dB with a standard deviation of 3.8. Note that there is considerable variation from speaker to speaker in the three phonation types; but for each speaker on this measure the value for creaky voice is less than that for modal voice, and the value for modal voice is less than that for breathy voice. This measure separates out breathy voice successfully in that for all speakers the value for breathy voice is higher than that for modal voice for any speaker. But it is less successful in distinguishing between modal voice and creaky voice in terms of their absolute values; on this measure speaker three's modal voice has an absolute value less than speaker four's creaky voice. It is only in relative terms that each speaker markedly differentiates the three phonation types on this measure.

Waveforms

Yet another useful display provided by the digital sound spectrograph is that of the waveform. As noted earlier, the wide band spectrograms show that creaky vowels have speech jitter (i.e. irregularly spaced pulses). In the waveform displays these irregular intervals are more clearly evident than in the wide band spectrograms.

Displayed in Figure 4 are the first 105 msec of the creaky vowels for all speakers, and for comparison the modal and breathy vowel displays for speaker five. The breathy vowel is characterized by an onset of indiscernible pulses; the modal vowel has regular pulses. Our measure of the jitter, is the variation in the interval between adjacent pulses. It can be seen in these displays that the onset of each creaky vowel pulse is clearly discernible for each of the speakers except for speaker three. For this speaker, in the span from 15 to 60 msecs there are several possible interpretations of what constitutes a pulse. Since our hypothesis is that creaky vowels are characterized by irregularly spaced pulses, we interpreted the distance between pulses in such a fashion as to produce the greatest regularity. Thus if irregularity persists, then it is present in the face of the best counter-claim interpretation against the hypothesis. The small arrows above speaker three's waveform indicate the points that were used in our measurements.

Table 1 presents a summary of the variance between pulses for creaky vowels and modal vowels in Jalapa Mazatec. A comparison of the two columns shows that the modal vowels for each speaker are characterized by uniform distances between pulses; the mean variance (.08 msecs) is small. On the other hand, creaky vowels vary widely in distances between pulses and have a large mean variance (9.1 msecs). Although there is considerable variation in the degree of creaky voice

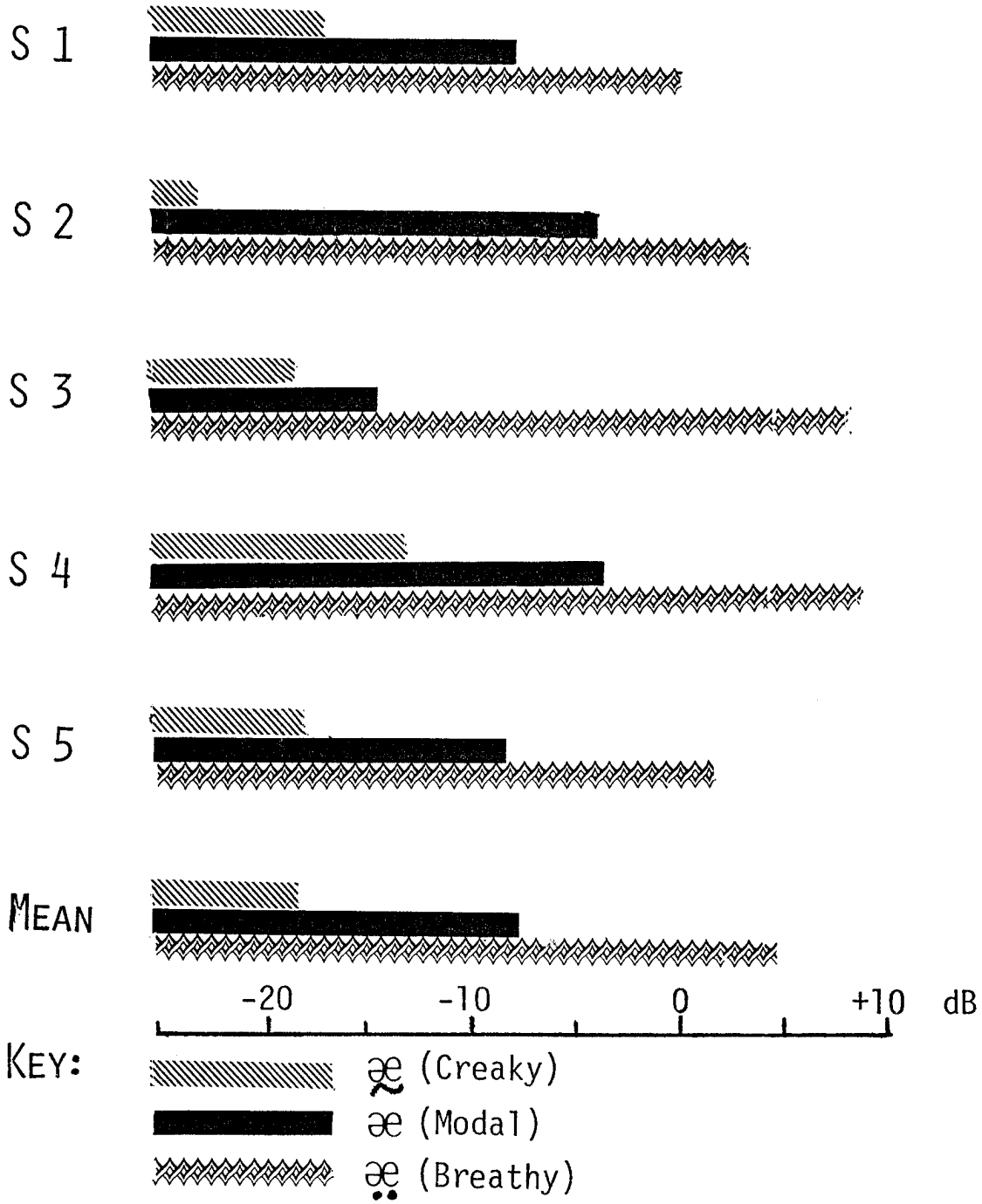


Figure 3: Relationship of the fundamental to the first formant in Jalapa Mazatec vowels.

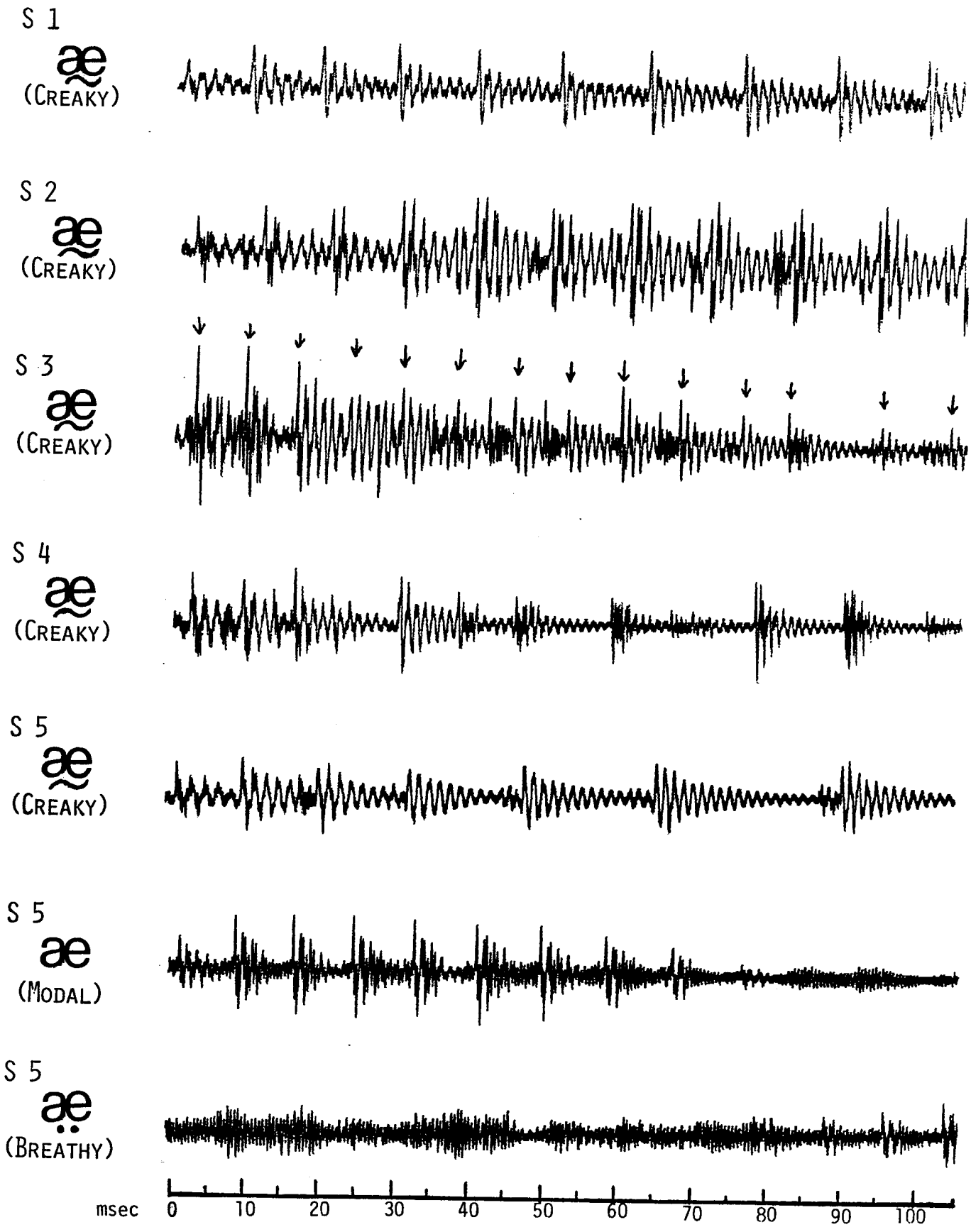


Figure 4: Waveforms of Jalapa Mazatec creaky vowels for five speakers and modal and breathy vowels for the fifth speaker.

among speakers, nevertheless for all speakers the value for creaky voice is higher than that for modal voice for any speaker. This measure, therefore, provides a good way of separating these two phonation types in absolute terms.

Table 1: Jitter between pulses

	S1	S2	S3	S4	S5	Mean	s.d.
(Creaky)	.39	.34	4.73	25.72	14.50	9.10	10.90
(Modal)	.04	.08	.02	.15	.11	.08	.05

Conclusion

Wide band spectrograms, narrow band power spectra, and waveform displays are clearly useful for analyzing voice qualities; they provide measures that can be used to identify voice quality differences. The results also support the conclusion that to some extent linguistic differences involving voice quality may be a relative matter. Measured in terms of the intensity of the fundamental in comparison with the intensity of the first formant, what may be modal voice for one speaker may count as creaky voice for another. But in this respect voice quality is similar to other phonetic features such as vowel height: measured in terms of frequency of the first formant, what may count as a high vowel [I] for some speakers may count as a mid vowel [ɛ] for others. Furthermore on the basis of our Jalapa Mazatec data, it seems that there are some measures that can be applied to data produced by a digital sound spectrograph that will separate out, in absolute terms, creaky, modal, and breathy voice. Further research is needed to show whether these measures remain valid when applied to a larger number of speakers, and other languages.

Acknowledgments

A version of this paper was presented at the 106th Meeting of the Acoustical Society of America (November 7-11, 1983). We gratefully acknowledge the support for our work of USPHS grant NS 18163-02. Also Terence and Judith Schram provided assistance in the field during the time that the data for this paper was collected and we wish to thank them for clarifying and deepening our understanding of Mazatec through many happy hours of discussion.

References

Davis, S. (1976). Computer evaluation of laryngeal pathology based on inverse filtering of speech (Speech Communications Research Laboratory, Santa Barbara).

Fant, G. (1956). "On the predictability of formant levels and spectrum envelopes from formant frequencies" in M.Halle (ed) For Roman Jakobson, (Mouton, The Hague). 109-20.

Fant, G. (1973). Speech Sounds and Features. (MIT Press, Cambridge).

Gudschinsky, S. C. (1959). "Proto-Popolocan: A comparative study of Popolocan and Mixtecan," (Indiana University Publications in anthropology and linguistics, no. 5; Waverly Press, Baltimore, Md.).

Jamieson, A. R. (1977a). "Chiquihuitlan Mazatec phonology," in William R. Merrifield (ed) Studies in Otomanguean phonology. (Summer Institute of Linguistics, University of Texas, Arlington). 93-105.

Jamieson, A. R. (1977b). "Chiquihuitlan Mazatec Tone," in William R. Merrifield (ed) Studies in Otomanguean phonology. (Summer Institute of Linguistics, University of Texas, Arlington). 107-35.

Kirk, P. L. (1966). "Proto-Mazatec phonology," Ph.D. dissertation, University of Washington.

Kirk, P. L. (1970). "Dialect intelligibility testing: The Mazatec study," Internat. J. Amer. Ling. 36, 205-211.

Ladefoged, P. (1983). "The linguistic use of different phonation types," in D. Bless and J. Abbs, (eds) Vocal fold physiology: contemporary research and clinical issues (College-Hill Press, San Diego).

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

Pike, E. V. (1956). "Tonally differentiated allormorphs in Soyaltepec Mazatec," Internat. J. Amer. Ling. 22, 57-71.

Pike, K. L. and Pike, E. V. (1947). "Immediate constituents of Mazateco syllables," Internat. J. Amer. Ling. 13: 78-91.

Rensch, C. R. (1976). "Comparative Otomanguean phonology," Language Science Monograph 14 (Indiana University, Bloomington).

Schram, J. L. and Pike, E. V. (1978). "Vowel fusion in Mazatec of Jalapa de Diaz," Internat. J. Amer. Ling. 44, 257-61.

Recognition of Famous Voices Forwards and Backwards

Diana Van Lancker, Jody Kreiman, and Karen Emmorey

Paper presented at the 106th meeting of the
Acoustical Society of America

The considerable amount of work on voice identification and recognition has focussed on unfamiliar voices, with a few exceptions that have used as stimuli voices familiar to the listeners. These studies have suggested that familiar voices can be recognized relatively easily from short voice samples without benefit of contextual cues. Even though cerebral processing underlying recognition of familiar-intimate as compared with familiar-famous voices may differ, we elected to use famous voices as stimuli to be able to compare familiar voice recognition in large numbers of normal and brain-damaged subjects.

Recognition of familiar voices in general deserves special attention, we believe, for several reasons. It is a major natural human ability practiced daily; evidence from developmental studies has shown that recognition of familiar stimuli (both faces and voices) follows a different maturational schedule from ability to discriminate between the same kinds of stimuli when they are unfamiliar; and research on brain-damaged subjects has shown that in the case of faces, a familiar stimulus engages different cerebral mechanisms from similar unfamiliar ones.

Our interest has been in establishing facts about familiar voice recognition in the normal adult population. What percentage of known voices would subjects recognize, and from what kinds of samples? To investigate these and similar questions we undertook a series of experiments, the first of which we are reporting here.

Stimuli

From an original set of 64 voices of male entertainers, politicians, and others well-known in film, radio, and television, 45 were prepared for listening tasks. Brief excerpts (4-6 seconds in length) of each speech sample were selected to be free of any sort of speech trademarks, background noises, and identifying sounds. Multiple-choice answer sheets were prepared; foils for each target item were carefully chosen to be plausible sources for that sample with respect to rhetorical style, the topic or content of the sample, and voice characteristics. We wished our task to reflect true recognition as closely as possible, and so, when designing the multiple- choice portion of the experiment, we paid considerable attention to detail in order to reduce the opportunity for subjects to employ selection strategies extraneous to voice recognition.

The voice samples (45 test items and 5 practice items) were digitized and edited to eliminate long pauses and to create 4 second samples, divided into two 2-second portions. No words were repeated across segments.

Listening Tasks

Three listening tapes were prepared. 1) 50 2-second samples in an unlimited set task; 2) 50 different 2-second samples, each with 6 possible answers; and 3) 4-second samples presented backwards, each also with 6 possible answers.

Task 1 tested subjects' abilities to recognize a famous voice from an unlimited set and without verbal or situational context. Subjects heard a 2-second sample over a loud speaker, and rated the familiarity of the voice on a 1-5 scale. The name of the speaker was then projected onto a screen, after which each subject answered the question: "Is it who I thought it was?" Our purpose in arranging Task 1 in this way was to probe voice recognition without requiring name retrieval.

In Task 2, subjects heard the second set of 2-second samples in a different order and with 6 names to chose among. After hearing each voice sample, subjects circled a name.

In the third listening task, subjects heard the entire 4-second sample for each target item played backwards. The test items were rerandomized and refoiled for this task.

Between listening Tasks 2 and 3, subjects filled out information sheets giving personal history and tv and movie-watching habits. At the end of the experiment, they filled out questionnaires indicating which voices they felt they would normally recognize.

Subjects

Since familiarity with famous persons varies somewhat with age, we divided our subjects into 3 age groups: 25 and under, 25-45, and over 45. A fourth group of subjects heard the backwards stimuli first. We will return to this 4th group in a moment. A total of 94 subjects were tested.

Results

Since our intent was to study recognition of familiar voices, the analyses reported here include only the voices subjects reported they knew and would normally recognize.

Results averaged across the first 3 groups are shown in Figure 1. For Task 1 (recognition of voices from an unlimited response set), the mean recognition rate for the three groups was 26.6 % of known voices. The three groups overall differed very little on this task, although individual subjects did vary --more on this task than on the other two. The average percent correct for listening task 2 (2 seconds forward, 6 choices) was 71.4 % correct; and for Task 3 (4 seconds backwards) the mean recognition rate was 59% . These values for tasks 2 and 3 are considerably above chance (16.7%), indicating that subjects are relatively successful at recognizing voices of famous persons given short, context-free samples presented forwards or backwards.

To insure that our subjects in Groups I-III were not practiced with respect to our voice ensemble by the time they reached the backwards presentation in Task 3, we tested another group of subjects (Group IV) who heard the backwards stimuli first. These subjects did not receive Task 1, and had no priming or other indication of the identities of our target set of speakers.

Group IV achieved a recognition rate of 52.2% on the backwards voices, and 64.7% on the forwards task (which was presented to them second). These values compare quite well with those for the other groups (Figure 2); and all values are well above chance. In addition, the decrement in performance from

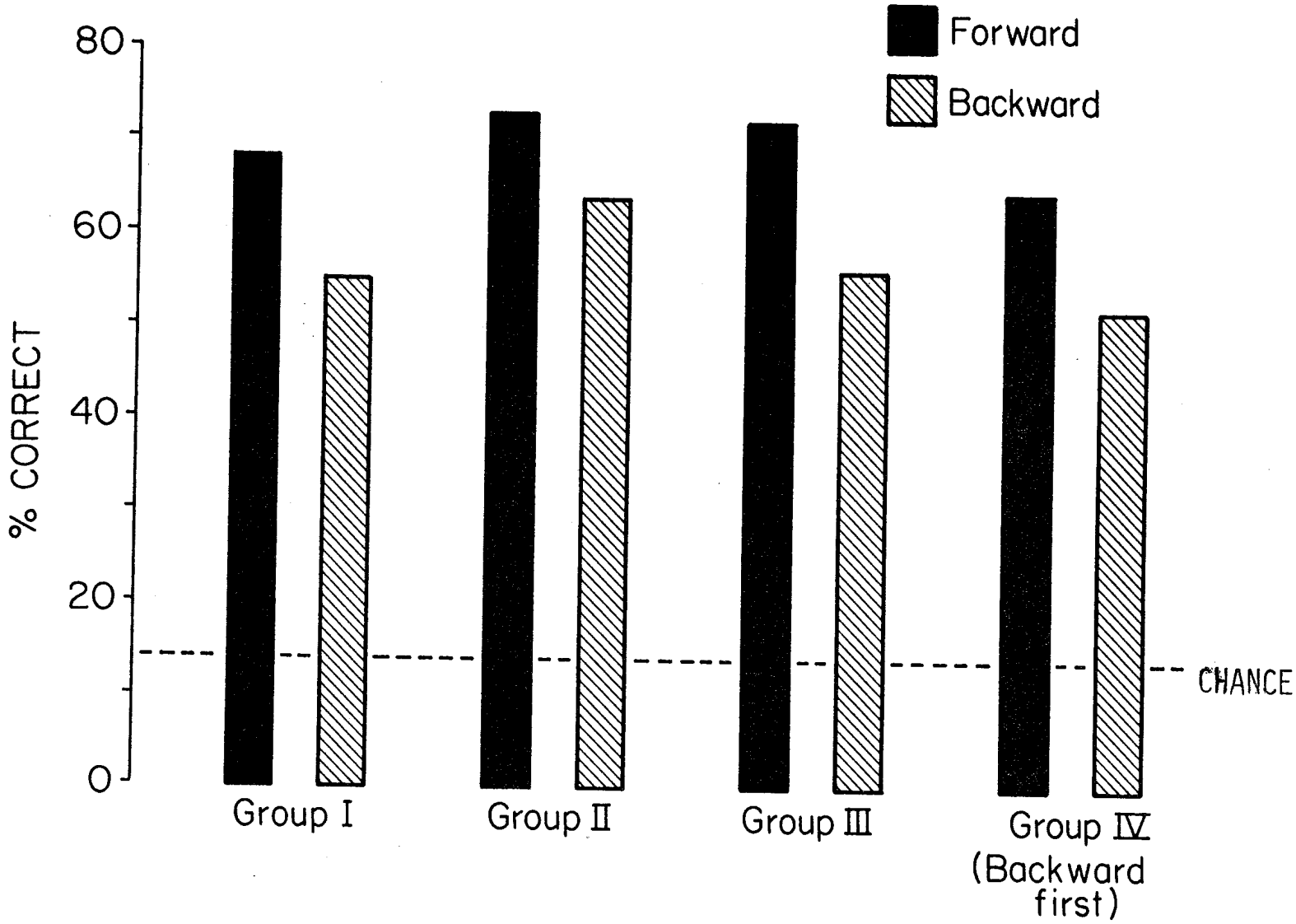


FIGURE 2. PERCENT CORRECT RECOGNITION OF FAMOUS VOICES ON FORWARD AND BACKWARD PRESENTATIONS, FOUR GROUPS.

Voice	% Correct Forward	% Correct Backward	Difference
Steve Martin	87.1%	80.6%	6.5%
W.C. Fields	95.5%	87.6%	7.9%
Vincent Price	79.0%	82.7%	-3.7%
Jack Benny	96.3%	85.4%	10.9%
Lawrence Welk	86.8%	36.8%	50.0%
David Frost	91.3%	41.3%	50.0%
Bob Hope	78.0%	28.6%	49.4%

TABLE 1. EXAMPLES OF SCORES ON FORWARDS COMPARED WITH BACKWARDS PRESENTATIONS OF FAMOUS VOICES.

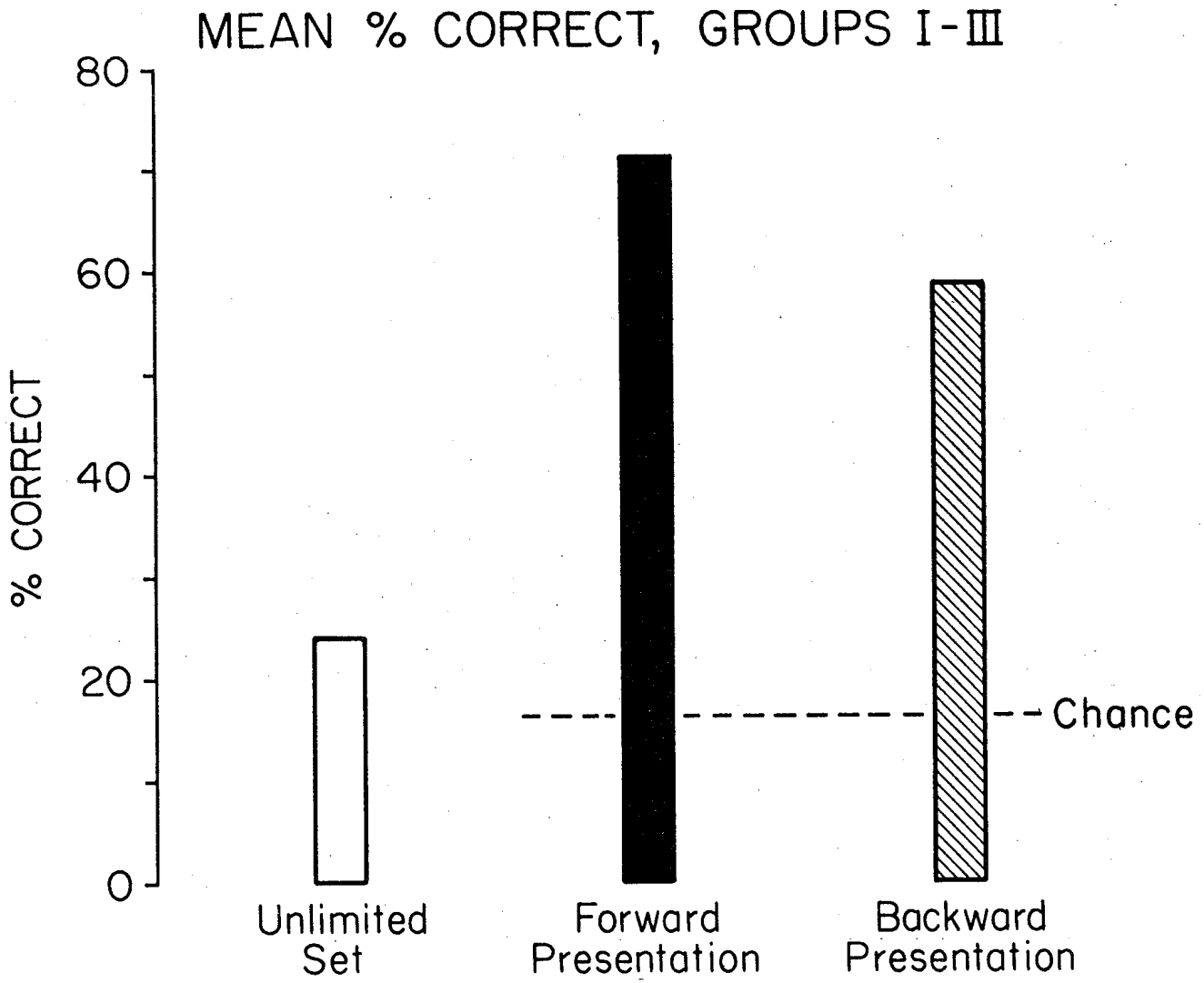


FIGURE 1. PERCENT CORRECT RECOGNITION OF FAMOUS VOICES ON THREE TASKS.

forwards to backwards presentation was remarkably stable across the 4 groups. The average decrement in performance for the first 3 groups was 12.8%; for group IV, the figure is 12.5%. Since we cannot attribute the success of Group IV in identifying backwards voices to familiarity with the target set, we infer that this was not a major factor affecting the performance of groups I-III, since performance levels in the four groups are quite comparable.

Furthermore, a 2-way analysis of variance comparing the 4 groups on Tasks 2 and 3 revealed significant effects of group and task, but no group by task interaction. Thus differences in performance between forwards and backwards presentation were not significantly different in the 4 groups.

It is interesting to note that errors were not evenly distributed across the voices (Table 1). Some voices were recognized nearly equally well forwards and backwards, (for example, Steve Martin, W.C.Fields, Vincent Price, Jack Benny), while some (for example, Lawrence Welk, David Frost, and Bob Hope) showed large decrements in recognizability when played backwards.

Subjects in this experiment were reasonably good at recognizing famous voices, given very short voice samples and, in the backwards condition, somewhat limited information about the voice. We are currently investigating the role of various acoustic parameters in successful recognition of individual famous voices; for the time being, we can conclude from performance on backward voices that voice recognition can be achieved from acoustic information limited to rate, pitch, pitch range, voice quality, and perhaps vowel quality, but without benefit of acoustic detail reflecting articulatory and phonetic patterns.

RECOGNITION OF FAMOUS VOICES GIVEN VOWELS, WORDS, AND TWO-SECOND TEXTS

Karen Emmorey, Diana Van Lancker, and Jody Kreiman

Paper Presented at the 106th Meeting of the
Acoustical Society of America

The experiment described here looks further at famous voice recognition using methods similar to those described in the preceding paper (Van Lancker, Kreiman, and Emmorey). We wanted to assess familiar speaker recognition abilities given different amounts and types of voice information. Since vowels constitute most of the acoustic energy in speech, it seemed logical to investigate how much they contribute to recognition of famous voices. If vowels are excerpted or isolated from the speech stream, will they still provide sufficient information for speaker identification? Previous research has suggested that a greater variety of phonetic features is correlated with better voice recognition. This experiment examined subjects' ability to identify famous voices given three different kinds of voice samples: two-second connected texts, single words, and vowel strings.

25 famous voices were chosen for this experiment from the voices used in the preceding study. One of the two second samples was selected for each voice. These samples made up the first condition.

A second tape was made containing a single word, unique to each speaker, and not appearing in the two-second sample. The mean length of the words was 481 msec. Each word was recorded twice with an interval of two-seconds between tokens; this was done to provide an opportunity for the subject to orient to the task and process the signal, while limiting the voice information to one word. These samples constituted the "single-word condition."

For the third condition, vowels were excerpted and concatenated without pauses. Like the words, the vowel string for each item was recorded twice with a two-second interval between. The number of vowels which made up the string for a particular famous person ranged from two to five, and the mean length of a single excerpted vowel was 142 msec. The mean length of a vowel string was 494 msec. This set of stimuli comprised the "vowel condition". The word and vowel stimuli for each voice were matched for length, in that each vowel string was within 50 msecs of the length of the word for that same voice.

As in the previous experiment, multiple choice answer sheets were prepared for each condition. Subjects were given six choices and asked to provide confidence ratings for each response. Subjects were not told that the target voices in all conditions were the same, but they were told at the beginning of each condition what type of stimuli they would hear. At the end of the listening tasks, the subjects filled out a questionnaire indicating whether they thought they would normally recognize each voice.

40 subjects, aged 16 - 64, volunteered to participate in the experiment. Subjects were divided into two groups of 20. One group received the listening tasks in the order: two-second texts, words, vowels. For the other group the reverse order was presented.

The results are as follows: first, we included measures only on target voices which each subject indicated that he would recognize. Performance was

significantly above chance on all three tasks as shown by Figure 1. The mean percent correct for two-second texts was 61%, for words, 40%, and for vowels, 34%. A two-way repeated measures ANOVA revealed a main effect of condition but no effect of presentation order. Subjects performed significantly better on the two-second text condition than on the word or vowels condition. The word and vowels conditions did not differ significantly from each other.

To test for the influence of guessing strategies on responses, the percent correct of unknown voices (voices which subjects said they would not recognize) was calculated. The mean percent correct for two-second texts was 26%, for words 24%, and for vowels 9%. Because these scores were significantly different from the scores for known voices, we take this as evidence that subjects' relatively good performance on these tasks was not due to some extraneous guessing stragey.

For our analysis of confidence ratings, we included all voices (known and unknown), since we wanted to examine confidence ratings for both familiar and unfamiliar voices. As would be expected, subjects were more confident of correct answers than of incorrect answers. Furthermore, they were more confident of their responses in the two-second text condition than in the words or vowels condition. They were about equally confident in the word and vowel conditions. This is further evidence that there was infact no difference in subjects' performance on the word and vowel recognition tasks.

The fact that there was no difference between these conditions must be interpreted in the context of what has been claimed previously for familiar voice recognition. Bricker and Pruzansky (1966) found that identification accuracy improved with the number of phonemes. Our data followed this trend, but the differences observed were not significant.

The difference between the subjects performance in our experiment and Bricker and Pruzansky's may be due to a difference in recognition ability for familiar-intimate versus familiar-famous voices. Bricker and Pruzansky used voices of people known personally by the listeners. The strategies and cues used to recognize intimate voices may be quite different from those used to recognize famous voices.

We thought it might be possible that within one stimulus type the number of phonemes would show a correlation with percent correct. However, this turned out not to be the case, as shown in Figure 2. There was no apparent relation between percent correct and the number of phonemes in the word or vowel string samples. The percentage correct does not rise with the number of phonemes.

It was further hypothesized that the number of distinct phoneme types would be a factor. This hypothesis was only relevant for the vowel stimuli because some types of phonemes were repeated within a number of vowel strings; but for the words, a phoneme was rarely repeated within the same word. The number of different phonemes also turned out to be poorly correlated with percent correct.

The overriding factor which seems to influence our results then is duration, not number or type of phoneme. For a particular voice the number and type of phoneme varied between the word and vowels stimuli, but this had no apparent effect on recognition. That is, the stimuli with more phonemes were not better recognized than ones with less. Duration and voice recognition ability remained constant for the words and vowels. Voice recognition, of course, increased with the two-second texts.

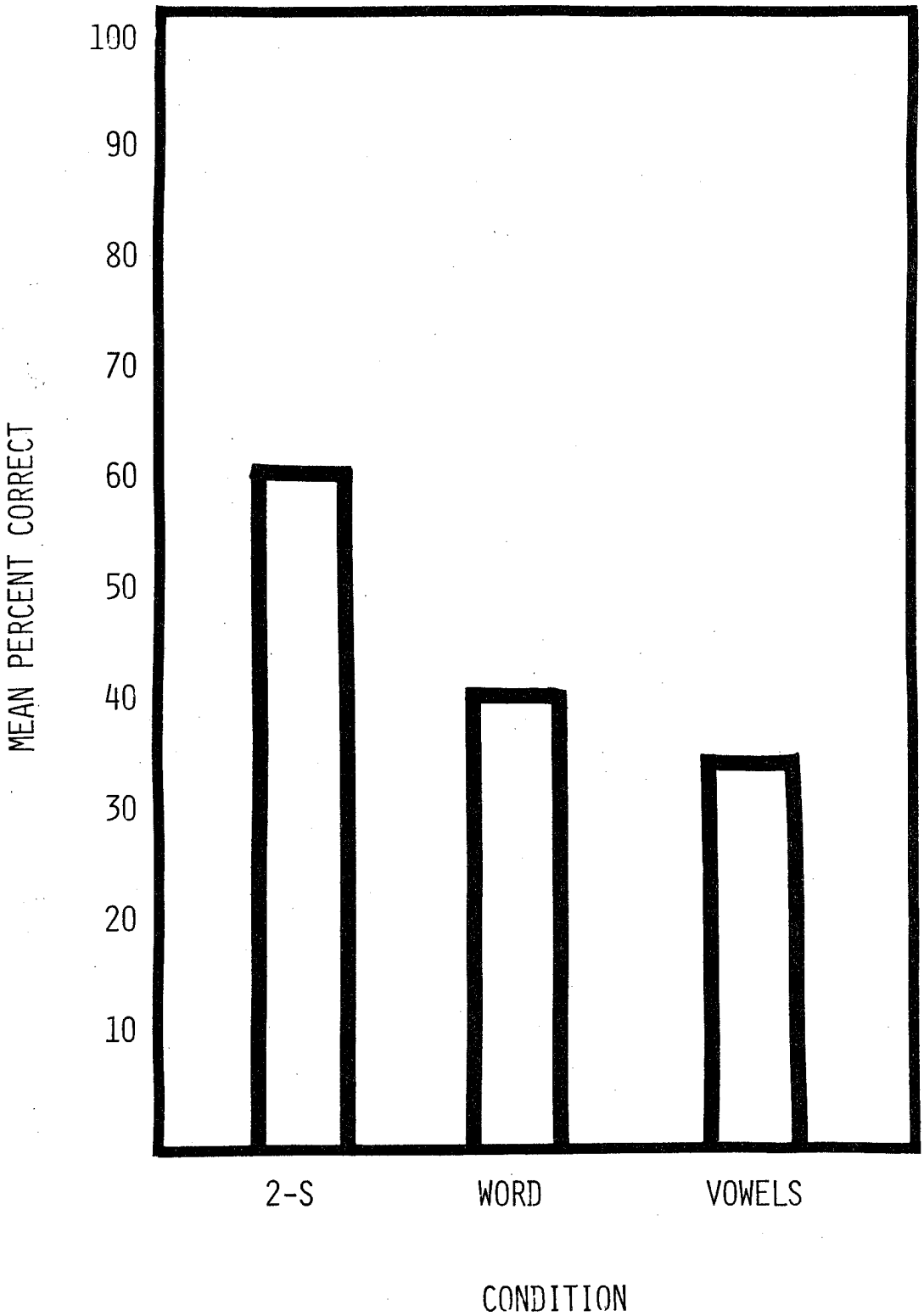
One question we originally sought to answer was whether vowels isolated from the speech stream provide sufficient information for speaker recognition. Our results indicate that vowels alone do provide sufficient information for recognition, but with less than 50% accuracy. The phonetic information carried by vowels in continuous speech appears to be approximately equal to that of a word with similar duration.

Topics for further investigation may concern individual voice differences. We noted that some voices were recognized much more easily given the vowel stimuli than the word and vice versa. Also, the contributions of duration, phoneme number, and phoneme type to voice recognition needs to be investigated in more detail.

Reference

Bricker, P. and Pruzansky, S. (1966) Effects of stimulus content and duration of talker identification. Journal of the Acoustical Society, 40, 1441-1449.

FIG. 1


100
90
80
70
60
50
40
30
20
10
MEAN PERCENT CORRECT
2-S
WORD
VOWELS
CONDITION