6.541J Handout 2/26/91

Journal of Phonetics (1989) **17,** 3–45

On the quantal nature of speech

Kenneth N. Stevens

Research Laboratory of Electronics and Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge MA 02139, U.S.A.

When a parameter specifying the configuration or state of an articulatory structure is manipulated through a range of values, some acoustic parameter describing the resulting sound often changes in a non-monotonic fashion. In particular, there appear to be ranges of the articulatory parameter for which there is very little change in the acoustic parameter and other ranges where the acoustic parameter is more sensitive to changes in articulation. The articulatory–acoustic relations are quantal in the sense that the acoustic patterns shows a change from one state to another as the articulatory parameter is varied through a range of values. Similar kinds of relations are observed between parameters that give a measure of the auditory response to a speechlike sound and parameters that specify some acoustic dimension of the sound. In this paper a number of examples of these types of acoustic–articulatory or auditory–acoustic relations are given. It is suggested that this tendency for quantal relations among these acoustic, auditory, and articulatory parameters is a principal factor shaping the inventory of acoustic and articulatory attributes that are used to signal distinctions in language.

1. Introduction

In this paper we shall explore certain properties of the human vocal tract as a generator of sound and the human auditory system as a receiver of the types of sounds that occur in speech. In particular, we want to examine the nature of two kinds of relations among the levels of description of speech events: (1) relations between vocal-tract configurations or states and the properties of the sound that results from these articulations, and (2) relations between acoustic parameters of the type that are observed in speech and auditory responses to sound described by these parameters. As a consequence of this survey, we shall make some observations about the role of articulatory and auditory constraints in shaping the inventory of phonetic features that are used distinctively in language.

In our investigation of sound generation, we propose to examine how certain acoustic parameters change as the various structures that form the vocal tract are manipulated through ranges of states and configurations. We will give a number of examples for which the relation between an acoustic parameter that can be observed in the sound and an articulatory parameter that can be manipulated by a speaker takes a particular non-monotonic form. In these examples, there are certain ranges of the articulatory

0095–4470/89/010003 + 43 $03.00/0

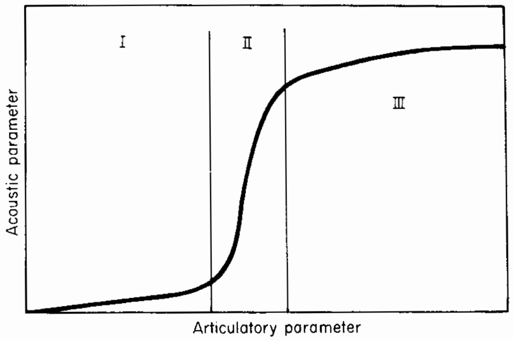

Figure 1. Schematization of a relation showing a change in a relevant acoustic parameter as an articulatory parameter specifying some aspect of the state or configuration of the speech-production system is manipulated. The curve can be divided into the regions I, II, and III as discussed in the text. The same form of the function can occur when an auditory parameter is plotted on the ordinate and an acoustic parameter is plotted on the abscissa.

parameter within which the acoustic parameter is quite sensitive to changes in the articulation. That is, the acoustic parameter undergoes large changes for relatively small manipulations of the articulatory parameter. Within other ranges of values for the articulatory parameter, the acoustic attribute remains relatively insensitive to perturbations in the articulatory parameter.

The situation is represented schematically in Fig. 1, which shows a hypothetical relation between some acoustic parameter in the sound radiated from the vocal tract and some articulatory parameter that takes on a series of values as indicated on the abscissa. The region where there are large changes in the acoustics for small shifts in articulation is designated as II in the figure. In regions I and III there is a plateau in the curve, or at least the articulatory–acoustic relation is not as steep, indicating that the acoustic parameter remains relatively stable when small modifications are made in the articulation. The difference in the value of the acoustic parameter from region I to region III is large. That is, there is a significant acoustic contrast between these two regions, which are separated by the intermediate region II in which there is a rather abrupt change in the acoustic parameter. As we will see in the examples, the difference in the acoustic pattern between regions I and III should not be regarded as simply a matter of identifying two points on a scale of some acoustic parameter. Rather, the acoustic attribute often undergoes a qualitative change as the articulatory parameter moves through region II.

Our review of some data on the processing of speechlike sounds by human listeners will be illustrated by a number of examples for which there is evidence for a relation of the form shown in Fig. 1. In this case, however, the ordinate on the graph is some parameter of the auditory response and the abscissa is an acoustic parameter of the type that can be under the control of a speaker. The examples will suggest that the measure of the auditory response, obtained through some psychophysical procedure, or, in some cases, through electrophysiological methods, shows a non-monotonic change as an acoustic parameter is manipulated. Region II can, in some sense, be considered as a threshold region such that as the acoustic parameter changes through this region the auditory response shifts from one type of pattern to another.

A conclusion to be drawn from these examples is that there are some articulatory states or configurations or gestures that give rise to well-defined patterns of auditory response in a human listener, such that these patterns are not strongly sensitive to small perturbations or inaccuracies in the articulation. These patterns are distinctive in the sense that if some articulatory parameter crosses over a threshold region there will be a significant change or a qualitative shift in the auditory response. The multidimensional space that depicts acoustic–articulatory or auditory–acoustic relations, rather than showing continuous and monotonic variation, exhibits quantal attributes characterized by rapid changes in state over some regions and less abrupt variations or greater stability over other regions.

We suggest that this tendency for quantal relations between articulatory and acoustic parameters or between acoustic and auditory parameters is a principal factor shaping the inventory of articulatory states or gestures and their acoustic consequences that are used to signal distinctions in language. The articulatory and acoustic attributes that occur within the plateau-like regions of the relations are, in effect, the correlates of the distinctive features. One component of the underlying representation of an utterance is in terms of these features, represented as some kind of matrix, possibly in a hierarchical form. Articulatory movements occur whenever a change in one or more features is specified by this input representation. Articulatory–acoustic or acoustic–articulatory relations of the type shown in Fig. 1 have several consequences for the nature of the sound pattern that emerges when a feature change is implemented. One consequence is that during the time the articulatory structures are close to the target state specified by a particular feature, some change in this configuration or state can occur without a significant modification in the relevant attribute of the sound pattern. Thus as the articulatory state undergoes a continuous sequence of maneuvers toward and away from the target value, the acoustic parameter resulting from this articulatory gesture may remain relatively stable over some part of this sequence. Furthermore, the precision with which the target articulatory state is achieved may be rather lax. As the articulatory parameter passes through values in region II of Fig. 1 when a feature change is implemented, there will tend to be a rapid change in the relevant acoustic parameter. This rapid change marks an event or a landmark in the acoustic stream. In the acoustic signal, therefore, there will be an alternation between temporal regions where the acoustic parameters remain relatively steady, and narrow regions marked by acoustic events where there are rapid changes. These somewhat discontinuous attributes of the acoustic signal occur in spite of rather continuous movements or changes in the articulatory parameters.

The articulatory–acoustic relation in Fig. 1 has been drawn to suggest that within region I or region III the acoustic parameter is not completely insensitive to changes in the articulatory parameter. Rather, the acoustic parameter in region III, for example, can show a greater contrast with the acoustic parameter in region I if the articulatory parameter achieves a more extreme value. The implication is that there can be differences in the strength with which a particular phonetic feature can be represented in the sound. There is evidence from acoustic analysis studies that the strength of the property used to implement a feature is dependent on the other features that co-occur with that feature (see, for example, Stevens, Keyser, & Kawasaki, 1986). These other features are often redundant and their implementation may be variable, and consequently they may enhance another feature by varying amounts. Furthermore, it has often been observed that the strength of the acoustic correlate of a particular feature may differ from language to language. In this paper we will not, however, dwell on the implications of this aspect of the acoustic–articulatory or auditory–acoustic relations.

We turn now to an examination of a number of examples of relations between acoustic and articulatory parameters and between auditory and acoustic parameters.[1]

2. Some acoustic–articulatory relations involving vocal-tract resonances

2.1. *Coupled resonators*

When the vocal tract is narrowed in some region along its length, the resulting configuration can often be viewed as a pair of coupled resonators. One resonator is the portion of the vocal tract behind the constriction, and the other is the portion in front of the constriction. The natural frequencies of the resonator combination are approximately equal to the natural frequencies of the individual resonators, with some perturbation from these values due to the acoustic coupling between the resonators. When the constriction is appropriately placed, the natural frequencies of the system can be relatively insensitive to changes in certain dimensions that specify the configuration of the vocal tract. A detailed analysis of vocal-tract acoustics, showing the natural frequencies for a variety of vocal-tract shapes, has been presented by Fant (1960). The relations between constriction position and formant frequencies for a particular model of the vocal-tract shape for vowels and consonants have also been examined by Stevens & House (1955, 1956).

To illustrate some of these properties of coupled resonators, we consider the simple configurations shown in Fig. 2. For one of these configurations, the tube is divided into two parts by a narrow section whose length is short compared with the length of the wider sections on either side of the combination. The other configuration consists of a relatively narrow uniform tube coupled to a tube with a much larger cross-sectional area.

We define a dimension l_1 in each case as the length of the left portion of the configuration that is closed at the left-hand end, and we manipulate the length l_1 while keeping the total length l of the tube constant. In the case of the configuration in Fig. 2(a), the

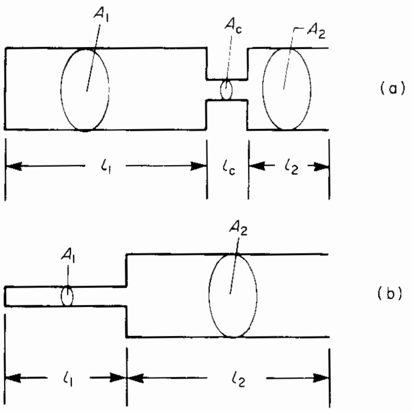

Figure 2. Examples of two configurations of acoustic tubes for which there is relatively little acoustic coupling between the left-hand portion, of length l_1, and the right-hand portion, of length l_2.

[1] Some of these relations were discussed in an earlier paper (Stevens, 1972). The discussion in the present paper represents a substantial extension and reformulation of the earlier material.

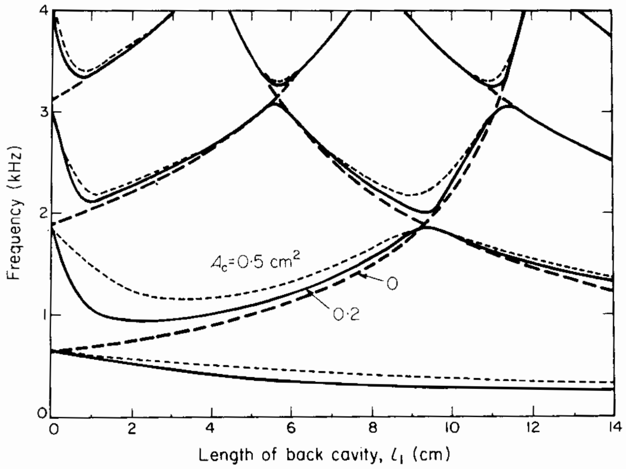

Figure 3. Frequencies of first four natural frequencies for configuration of Fig. 2(a), as the length l_1 of the back cavity is manipulated. The total length $l_1 + l_c + l_2 = 16$ cm, the constriction length $l_c = 2$ cm, and the cross-sectional areas A_1 and A_2 are 3 cm^2. The long-dashed lines (labeled 0) correspond to the case where $A_c \ll A_1$, in which case the lowest natural frequency is zero. The solid and short-dashed lines are for $A_c = 0.2$ and 0.5 cm^2, respectively. The tube is assumed to be hard-walled, and the radiation impedance is assumed to be zero.

length l_2 of the front cavity is $l_2 = l - l_1 - l_c$, where l_c is the length of the constriction, so that l_2 decreases as l_1 increases. The "uncoupled" natural frequencies of the posterior component are given by

$$0, \frac{c}{2l_1}, \frac{c}{l_1}, \frac{3c}{2l_1} \ldots$$

while the natural frequencies of the anterior component are

$$\frac{c}{4l_2}, \frac{3c}{4l_2}, \ldots,$$

where c = velocity of sound. These frequencies are plotted as a function of l_1 in Fig. 3, as dashed lines. We have selected a nominal overall length of 16 cm for purposes of illustration, and we have neglected the effects of radiation. These are the natural frequencies for the ideal case where the cross-sectional area A_c of the constriction is zero. There are several points of intersection where the uncoupled natural frequencies of the two component tubes are equal.

When the cross-sectional area A_c is non-zero, but is small compared with A_1 and A_2, the cross-sectional area of the larger tubes (we assume that $A_1 = A_2 = A$ in this example), there is some influence of one resonator on the other, and the natural frequencies of the coupled system are shifted slightly from the values for the uncoupled resonators. The lowest natural frequency, which is zero for the uncoupled case, becomes approximately equal to

$$\mathrm{F}_1 = \frac{c}{2\pi\sqrt{Al_1\left(\dfrac{l_c}{A_c} + \dfrac{l_2}{A}\right)}} \tag{1}$$

if we assume the resonator walls to be hard, and all the other natural frequencies tend to be shifted upwards. At a value of l_1 corresponding to a point of intersection for the uncoupled resonator, the natural frequencies are no longer equal, but are split apart as a consequence of acoustic coupling. One of the two frequencies is equal to the frequency at the original point of intersection, and the other is higher by an amount approximately equal to

$$\Delta F = \frac{c^2}{2\pi^2 l_c l_2 F_n} \times \frac{A_c}{A}, \tag{2}$$

where F_n is the uncoupled natural frequency. Equation (2) follows from simple acoustic theory of one-dimensional sound propagation in tubes. There is a limit to the proximity that can be achieved for two natural frequencies, depending on the area ratio A_c/A. For example, near the point $l_1 = 9.3$ cm in Fig. 3, the distance between F_2 and F_3 is about 200 Hz when $A_c = 0.2\,\mathrm{cm}^2$ and is about 400 Hz when $A_c = 0.5\,\mathrm{cm}^2$.

The natural frequencies of the coupled system of Fig. 2(a) are shown in Fig. 3 for two values of the area ratio A_c/A. In the vicinity of the values of l_1 corresponding to points of intersection for the uncoupled natural frequencies, the natural frequencies of the coupled system achieve maximum or minimum values, and these frequencies, therefore, are relatively insensitive to small changes in l_1 in the vicinity of these points. These regions are in the vicinity of $l_1 = 5.5, 9.3$, and 11.2 cm in the figure. There is also a region for l_1 in the range 2–4 cm where the Helmholtz resonance, given approximately by equation (1), is close to the lowest natural frequency of the front cavity. There is a broad minimum in F_2 in this region, and, although a maximum in F_1 is not evident in the figure, this formant is relatively insensitive to l_1 over this range.

The pattern in Fig. 3 is modified somewhat if the length l_c is greater, such that the lowest natural frequency of the constriction is within the range of the lower natural frequencies of the front and back cavities. If this constriction length remains fixed as the position of the constriction is varied, a constant natural frequency equal to $c/2l_c$ is added to the inventory of natural frequencies arising from the front and back cavities. An example of the modified pattern for $l_c = 6$ cm is given in Fig. 4. In this case, two sets

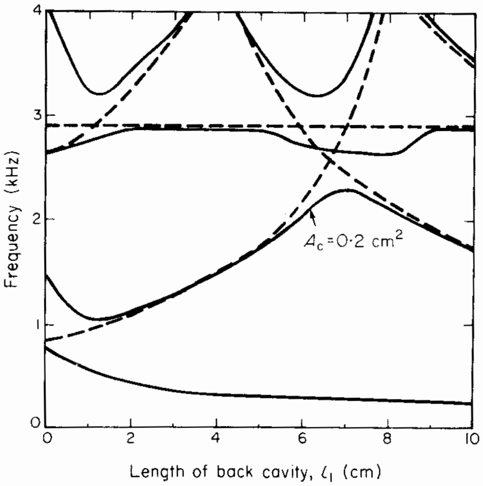

Figure 4. Same as Fig. 3, except that $l_c = 6$ cm, and frequencies are given only for $A_c \ll A_1$ (dashed line) and $A_c = 0.2\,\mathrm{cm}^2$ (solid line). In this case the natural frequency of the long constriction can be seen at about 2900 Hz.

of curves are shown—one for the uncoupled resonators (dashed lines), and one for a constriction size $A_c = 0.2\,\mathrm{cm}^2$. Again there are shifts of the frequencies from the values for the uncoupled resonators, particularly in the vicinity of the points of intersection. In this example, the frequency in the vicinity of 2900 Hz represents the natural frequency of the constriction. There is a region around $l_1 = 6.5$ cm where three uncoupled natural frequencies are very close together.

Similar patterns can be derived from the configuration in Fig. 2(b). The uncoupled natural frequencies for the two components of the system, with lengths l_1 and l_2, are given by

$$\frac{c}{4l_1}, \frac{3c}{4l_1}, \frac{5c}{4l_1}, \ldots$$

for the posterior section and

$$\frac{c}{4l_2}, \frac{3c}{4l_2}, \frac{5c}{4l_2}, \ldots$$

for the anterior section. These frequencies are plotted as a function of l_1 as dashed lines in Fig. 5, with the constraint that $l_1 + l_2 = l =$ constant. We observe points of intersection where the uncoupled natural frequencies of the two sections are equal. This pattern would be obtained if A_1 were extremely small compared with A_2.

Again the effect of acoustic coupling between the two tubes is to shift the natural frequencies away from their values for the uncoupled case. The modifications in the natural frequencies when coupling is taken into account are shown by the solid lines in Fig. 5. The principal effect is in the vicinity of the values of l_1 for which the natural frequencies of the uncoupled resonators are equal. The natural frequencies for the coupled system are no longer equal. If $A_1 \ll A_2$, it can be shown that the separation between the two

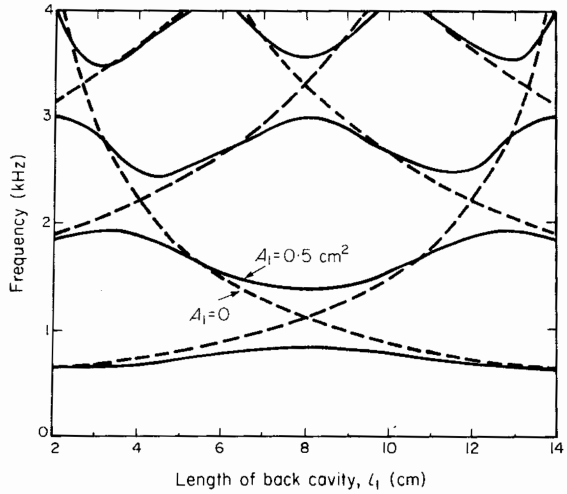

Figure 5. Frequencies of first four natural frequencies for configuration of Fig. 2(b), as the length l_1 of the back cavity is manipulated. The total length $l_1 + l_2 = 16$ cm, and the cross-sectional area $A_2 = 3\,\mathrm{cm}^2$. The long-dashed line corresponds to the case where $A_1 \ll A_2$, and the solid line is for $A_1 = 0.5\,\mathrm{cm}^2$. The radiation impedance is assumed to be zero.

natural frequencies is given by

$$\Delta F = \frac{2}{\pi} F \sqrt{\frac{A_1}{A_2}}, \tag{3}$$

where F is the uncoupled natural frequency. The relations between formant frequencies and the dimension l_1 for the case where $A_1/A_2 = 0.17$ in Fig. 5 again shows regions where the frequencies achieve maximum or minimum values. When l_1 is in one of these regions, such as in the vicinity of $l_1 = 8$ cm in the figure, the formant frequencies are relatively insensitive to changes in l_1, whereas between these regions (e.g., near $l_1 = 6$ cm), small perturbations in l_1 can give rise to substantial changes in the formant frequencies.

Figures 3–5 illustrate relations between resonator shapes and natural frequencies for a tube in which one part is substantially narrower than the rest of the tube. The three examples include cases where there is either a short or a longer constriction separating two wider portions of the tube, and a configuration for which a narrow section is coupled to a wide section. For all of the examples, the relations between constriction positions and formant frequencies show similar characteristics: constriction locations where particular formant frequencies show minimum or maximum values and are therefore rather insensitive to changes in constriction position, and other locations where the formant frequencies are more sensitive to perturbations in the constriction position.

The principles governing the relations between resonator shapes and natural frequencies in Figs 3–5 can be used as a guide in examining the relation between vocal-tract shapes and sound output as constrictions are produced at various points in the vocal tract by manipulating the tongue-body position and the lips. When speech sounds are produced with a source at the glottis, as is the case for vowels and sonorant consonants, the natural frequencies of the supraglottal system are manifested in the sound as peaks in the spectrum. In order to maintain sonorancy, the cross-sectional area of a constriction in the vocal tract should not be less than 0.2 to 0.3 cm^2. A narrower constriction would lead to a possible increase in intraoral pressure and to a modification of the glottal source, as a consequence of a pressure drop across the constriction. Since the average cross-sectional area along the vocal tract for an adult is about 3 cm^2 (corresponding to a vocal tract volume of roughly 50 cm^3 and a vocal-tract length of about 16 cm), then we can expect ratios of minimum to maximum cross-sectional area to be as small as 0.1 or less, i.e., a ratio encompassed in the examples in Figs 3–5.

In plotting the pattern of change of the natural frequencies for different resonator shapes in Figs 3–5 (and in the figures that follow), our principal concern is with natural frequencies extending up to about F$_4$, i.e., up to 3–4 kHz for the resonator length used here. There are two reasons for this limitation. One is that in the frequency range above 3–4 kHz, the auditory resolution, expressed in terms of the width of the critical bands, is poor. Consequently, listeners may be relatively insensitive to the detailed spacing of spectral prominences at these high frequencies. Secondly, the acoustic losses at these higher frequencies become greater, primarily as a consequence of increased radiation losses. As a result, the spectral peaks due to individual formants are not as prominent as they are at low frequencies, and changes in the frequencies of the peaks do not have as significant an effect on the overall spectrum shape as do similar changes in the frequencies of the lower formants.

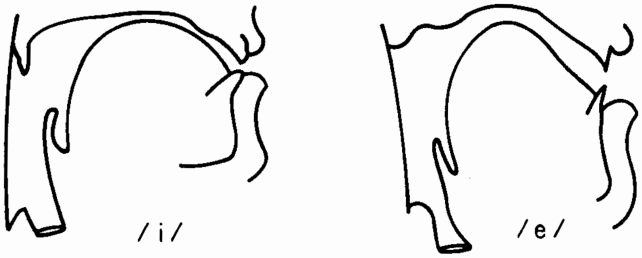

Figure 6. Examples of midsagittal vocal-tract configurations obtained from cineradiographs of utterances sampled in the nonlow front vowels /i/ (left) and /e/ (right). Adapted from Bothorel, Simon, Wioland & Zerling (1986).

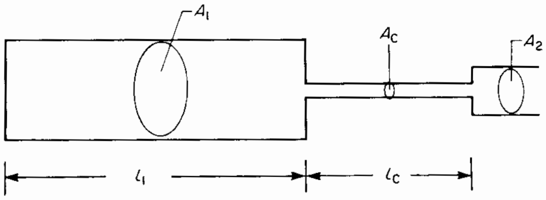

Figure 7. A resonator configuration approximating the vocal-tract area function for a non-low front vowel. Approximations for different non-low front vowels can be obtained by manipulating l_1 and l_c.

2.2. *Non-low front vowels*

A non-low front vowel is produced by displacing the tongue body forward to create a narrowing in the airway in the front portion of the vocal tract between the dorsum of the tongue and the hard palate. Two examples of midsagittal configurations for vowels of this type are shown in Fig. 6. These vowels are characterized by a wide pharyngeal region, occupying somewhat more than one-half of the total vocal-tract length, narrowing down to a relatively small cross-sectional area in the palatal region, with some widening at the teeth and lips. An approximation to the vocal-tract shape can be achieved with an idealized resonator configuration like that shown in Fig. 7. This shape is similar to that given in Fig. 2(a), except that the constriction length l_c is greater, and the cross-sectional area of the cavity in front of the constriction is somewhat less than that of the back cavity. Displacement of the tongue body in the anterior-posterior direction in Fig. 6 corresponds roughly to changing the length l_1 of the back cavity in Fig. 7.

Calculation of the first few natural frequencies of a hard-walled tube like that in Fig. 7 as l_1 is changed gives the result shown in Fig. 8. The two panels of the figure correspond to two different constriction lengths l_c. The main feature of interest in these patterns is a broad maximum of F_2 for configurations having a back-cavity length in the range 6.5 to 9 cm. In this region where F_2 is a maximum, this formant is relatively close to F_3. When the constriction is even farther forward, F_3 becomes close to F_4, while F_2 remains relatively high. The exact location of the maximum in F_2 and the distance between the formants in this cluster of F_2, F_3, and F_4 depend on the length and cross-sectional area of the constriction between the tongue dorsum and the hard palate. When the length of the back cavity decreases to the left of the F_2 maximum in Fig. 8, there is a substantial decrease in F_2, and F_2 becomes quite sensitive to changes in l_1. Proximity of F_2 and F_3 for some constriction position will always occur when a constriction is placed in the front

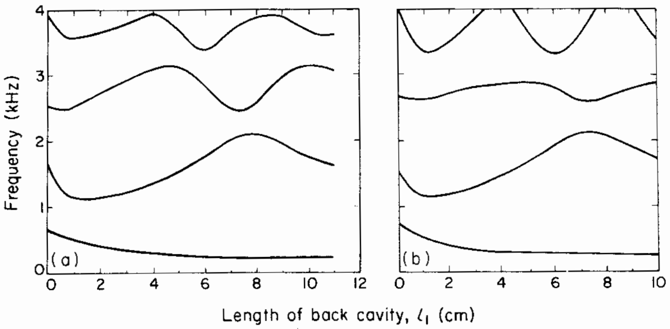

Figure 8. Natural frequencies of configuration of Fig. 7, as the length l_1 of the back cavity is manipulated. The two panels correspond to two different constriction lengths (a) $l_c = 5$ cm, and (b) $l_c = 6$ cm. The cross-sectional areas of the different parts of the tube are $A_1 = 3.0\,\text{cm}^2$, $A_c = 0.3\,\text{cm}^2$, and $A_2 = 1.0\,\text{cm}^2$, and the overall length of the tube is 16 cm. The portions of these curves for $l_1 > 6$ cm correspond approximately to the configurations for non-low front vowels.

part of the vocal tract. The simple configuration of Fig. 2(a) also shows such a region, as can be seen in Fig. 3.

We can conclude, then, that when the tongue body is in a fronted position such that a constriction is formed in the anterior part of the vocal tract, there is a range of positions for which F_2, together with F_3 and possibly F_4, are relatively insensitive to anterior–posterior perturbations in tongue-body position. Within this range of positions, F_2 is a maximum, and is within a few hundred Hertz of F_3. We note, however, that F_1 varies monotonically with constriction position and with constriction size for the configuration of Fig. 7. The sensitivity of the first formant to these parameters will be discussed later.

2.3. *Non-low back vowels: the effect of rounding*

We turn next to vowel configurations for which the constriction formed by manipulating the position of the tongue body is posterior to the palatal region discussed above. If we examine Fig. 3 for small values of l_1, i.e., corresponding to constrictions in the pharyngeal region, we observe a region where F_2 is relatively low and F_1 is high and close to F_2. In this region, these two formant frequencies are relatively stable in the sense that they are relatively insensitive to perturbations in the position of the construction. This pattern of formant frequencies corresponds to a low vowel, and will be discussed in more detail in the next section of this paper. Given the model of Fig. 2(a), there is apparently no region except near the glottal end of the vocal tract where the constriction can be placed to yield of a relatively low and stable value of F_2.

A minimum in F_2 as a function of constriction position in the upper pharyngeal region can, however, be achieved by increasing the acoustic mass at the anterior end of the vocal tract. The effect of this modification is to decrease the natural frequency of the front cavity in the configuration of Fig. 2(a). This increased acoustic impedance is realized by rounding the lips, which creates a section at the end of the vocal tract that is narrower than the cavity posterior to the lips. Midsagittal sections for two examples of vowels in which this strategy is used are shown in Fig. 9.

A configuration of resonators that can be used to explore the effect of constriction position and mouth opening on the natural frequencies is shown in Fig. 10. This

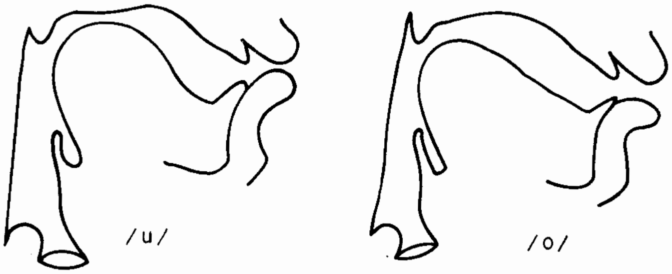

Figure 9. Examples of midsagittal vocal-tract configurations obtained from cineradiographs of utterances sampled in non-low back rounded vowels /u/ (left) and /o/ (right). Adapted from Bothorel *et al.* (1986).

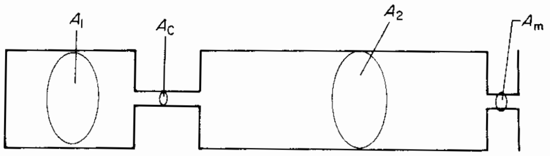

Figure 10. A resonator configuration approximating the vocal-tract area function for a non-low back rounded vowel.

configuration is similar to that in Fig. 2(a), except that there is some narrowing of the tube at the open end. The variation of the natural frequencies of this configuration as the constriction position is manipulated is given in Fig. 11. The figures shows a broad minimum for F_2 over a range of lengths of the back cavity. F_2 is within 100 Hz of its minimum value for l_1 between 2 and 7.5 cm. Within this range of l_1, the spacing between F_1 and F_2 is 400–500 Hz, and, while F_1 does not achieve a maximum value, it varies by only about 80 Hz. For a more realistic model that includes yielding walls, the variation of F_1 would be even less than this (Fant, 1972). The exact position of the constriction

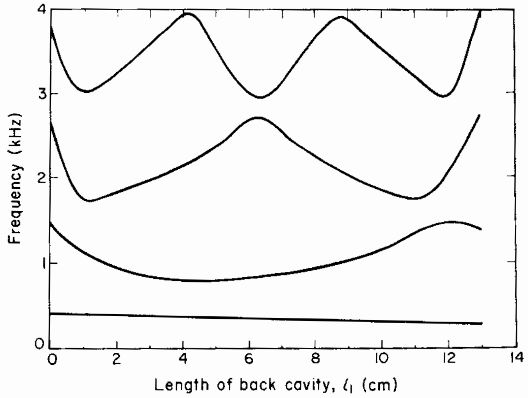

Figure 11. Natural frequencies of configuration of Fig. 10, as the length l_1 of the back cavity is manipulated. The cross-sectional areas of the different parts of the tube are $A_1 = A_2 = 3.0\,\text{cm}^2$, $A_c = 0.3\,\text{cm}^2$, and $A_m = 0.3\,\text{cm}^2$. The constriction lengths are 2 cm for A_c and 1 cm for A_m, and the overall length is 16 cm. The portion of this chart for l_1 in the range 2 to 6 cm corresponds approximately to the configuration for a non-low back rounded vowel.

for which a minimum of F_2 is reached depends upon the size of the opening at the radiating end of the tube and on the length and size of the constriction. In any event, for vocal-tract shapes that are rounded, like those in Fig. 9, there is a range of constriction positions that will yield a value of F_2 that is low, is relatively stable, and is close to F_1. These positions are achieved by manipulating the tongue body to a backed position to form a constriction in the region of the soft palate or the upper pharynx. For these configurations F_1 is lower than the values that could be reached if there were no constriction at the lips.

2.4 *Low vowels*

As we have seen in Figs 3, 4, and 5, and as can be predicted from acoustic theory, a relatively high value of the lowest natural frequency can be obtained by making a constriction in the posterior part of the vocal tract, and by forming a relatively large cross-sectional area in the anterior part of the tract. The large oral cavity is created by lowering the tongue body, and this lowering is usually facilitated by lowering the mandible.

Examples of midsagittal sections for low vowels are given in Fig. 12. These shapes illustrate the narrowing in the pharyngeal region, and the increase in cross dimensions at the transition into the upper pharynx or oral cavity.

The simple two-tube shape of Fig. 2(b) is a rough approximation to the configurations of Fig. 12, and Fig. 5 illustrates the pattern of change of the natural frequencies as the length of the narrow posterior portion is manipulated. Two regions of Fig. 5 are of interest in relating the two-tube resonator to possible configurations for low vowels. These regions are in the part of Fig. 5 up to a value of l_1 of about 9 cm. (We assume that it is not possible to adjust the tongue body to form a narrow tube that encompasses the entire pharynx plus a portion of the oral cavity, with an expansion to a larger cross-sectional area only in the anterior portion of the oral cavity.) For values of l_1 in the middle range of 7–9 cm, F_1 is a maximum and F_2 is a minimum, such that there is a minimum distance between these two formants. This shape is approximated by a backed and low tongue-body position, as in the left panel of Fig. 12.

In the vicinity of l_1 = 4 cm in Fig. 5, F_2 is close to a maximum value, and the separation between F_2 and F_3 becomes small. In this region, the value of F_1 is slightly less than its value for l_1 = 8 cm, but F_1 is still high (relative to its value for a uniform vocal

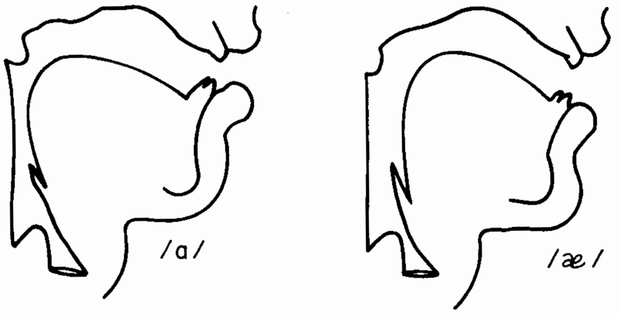

Figure 12. Examples of midsagittal vocal-tract configurations obtained from cineradiographs of utterances sampled in the low vowels /ɑ/ (left) and /æ/ (right) (from Perkell, 1969).

tract) and is only weakly sensitive to changes in l_1. This resonator shape can be approximated by placing the tongue body in a low and fronted position, while maintaining a relatively narrow constriction in the region of the tongue root or lower pharynx. Basically this shape can be viewed as a tube that is narrow in the lower pharynx and gradually widens or flares out toward the mouth opening.

2.5. *Effect of constriction size and rounding*

The above discussion has indicated that there are several ways in which a narrowing can be placed in some region along the length of the vocal tract such that one or more of the lowest two or three formants has a maximum or a minimum value. A consequence is that, when a constriction is located in the vicinity of one of these points, the frequencies of the formants are relatively insensitive to modifications in the position of the constriction. At other constriction locations, the formant frequencies are much more sensitive to changes in the constriction position. The examples we have given suggest that the front–back dimension and the low–non-low dimension for vowels can be based on these principles of formant proximity and coupled resonators: a maximum F_2 close to F_3 (and possibly F_3 close to F_4) for front vowels, F_2 close to F_1 for non-low back vowels that are rounded, and F_1 close to a maximum value for low vowels.

The relations between natural frequencies and resonator shapes discussed above show stable regions when the articulatory dimension on the abscissa defines the position of the constriction. When the cross-sectional area of the constriction is increased or decreased, keeping the constriction position fixed at one of the stable regions, the formants tend to change monotonically. That is, there is not a well-defined range of constriction sizes for which the formant frequencies achieve maximum or minimum values and thus are relatively insensitive to constriction size. However, the formant frequencies are usually not strongly sensitive to constriction size under most conditions.

For example, for the resonator shape in Fig. 2(b), and the corresponding pattern of natural frequencies in Fig. 5, we have seen that the spacing between the formants at the extreme points is approximately proportional to the square root of the area ratio. A change in area ratio of, say, 20% gives rise to a change in formant spacing of just 10%. Furthermore, in the case of resonator shapes with a low first formant, the mass of the vocal-tract walls modifies the formant frequency relative to its value predicted for a hard-walled resonator. The effect is to render this formant frequency less sensitive to the degree of constriction than would be predicted for a resonator with hard walls. Thus for high vowels, changes in constriction size brought about by manipulating the tongue-body height cause relatively small shifts in the first formant frequency.

Nevertheless, it is significant that articulatory manipulations that change degree of vocal-tract constriction do not give rise naturally to stable regions in the acoustic dimensions. If stability is to be achieved for the first formant frequency, it must be realized through appropriate shaping of the tongue surfaces as they make contact with opposing palatal surfaces. This method of obtaining stability in constriction size for a high front vowel [i] in the face of variable degree of contraction of the genioglossus muscle has been discussed by Kakita & Fujimura (1977). Further detailed study of anatomical and myoelastic constraints is needed to determine the extent to which similar non-linear relations between muscle excitation and displacement are available to assist a speaker in achieving particular degrees of tongue-body constriction for other vowels.

In addition to tongue height and backness, lip rounding can also be utilized to produce distinctions between vowels. As with degree of tongue-body constriction, manipulating the cross-sectional area and length of the lip opening gives rise to gradual and monotonic changes in the formant frequencies. That is, there appears to be no value of lip opening (or degree of rounding) leading to formant frequencies that show stable regions of the types observed in Figs 3, 4, and 5. However, it is not unreasonable to suppose that contractions of the muscles that are recruited to round the lips reach some kind of limit, beyond which it might be difficult to go without a significant increase in muscle excitation. That is, there are anatomical constraints that make it possible to reproduce a particular degree of rounding without requiring a great deal of precision in the degree of excitation of the appropriate muscles.

Rounding can have an important influence on the relations between formant frequencies and constriction position. In particular, presence or absence of rounding can shift the constriction locations where F_2 is a minimum and close to F_1 (corresponding to back vowels) and where F_2 is a maximum and close to F_3 (or F_3 is close to F_4). As we have already observed in Fig. 11, rounding for non-low back vowels makes it possible for F_2 to achieve a minimum value in a region where F_1 is relatively stable. When there is no lip rounding [as in Fig. 8(a), (b)], F_2 reaches a minimum, but the constriction position is close to the glottis, and F_1 is changing rapidly in this region. Without rounding, it is not possible to achieve a minimum value for F_2 when the constriction is in the middle or upper pharyngeal region. Put another way, when a constriction is located in the pharyngeal region, the first two formants are more sensitive to the position of the constriction when there is no rounding than when there is rounding. Rounding, furthermore, permits closer proximity of F_1 and F_2.

A different set of factors comes into play when rounding is imposed on front vowel configurations. Rounding of a front vowel will tend to lower the frequencies of the front-cavity resonances, and will also lower somewhat the resonance corresponding to the long constriction. Furthermore, because of the smaller mouth opening the acoustic losses due to radiation will be smaller and the bandwidths of the front-cavity resonances (and of the resonance corresponding to the long constriction) will be narrower. In particular, in a configuration like that in Fig. 7, which gives the pattern of natural frequencies in Fig. 8(a), and reproduced as the solid curves in Fig. 13, rounding causes a change in the formant curves to those shown by the dashed lines in the figure. The place where F_3 is a minimum and F_2 is a maximum is shifted to the right and the frequency is somewhat lower. The two formants are in closer proximity for the rounded configuration. The closer spacing of F_2 and F_3, together with the reduced radiation loss, will result in a spectral peak for these formants that is more prominent than it would be for the unrounded case. It appears, then, that rounding of non-low front vowels can lead to stable regions in which F_2 and F_3 are relatively insensitive to constriction position.

For both back and front vowels, then, one of the effects of rounding is to permit closer proximity of two formants—F_1 and F_2 in the case of back vowels and F_2 and F_3 in the case of front vowels. This proximity of a pair of formants creates a more prominent peak in the spectrum because of the mutual reinforcement of the contribution of these formants to the vocal-tract transfer function. Another factor contributing to the prominence or relative narrowness of the spectral peak is the tendency for one of the two closely-spaced formants (the higher frequency one) to have a considerably wider bandwidth than the other. This increased bandwidth arises because of the increased acoustic losses at the constriction separating the two resonators for the mode in which the sound pressure in

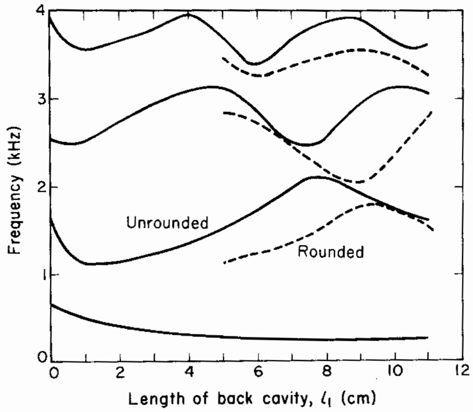

Figure 13. Showing the effect of rounding for a non-low front vowel. The solid lines are the natural frequencies of the "unrounded" configuration of Fig. 7, as already shown in Fig. 8(a). The dashed lines give the natural frequencies when rounding is imposed on the configuration of Fig. 7, by adding at the right-hand end a section with length 1 cm and cross-sectional area 0.5 cm^2.

the two resonators is 180° out of phase, i.e., for the asymmetric mode. For this mode there is a large volume velocity at the constriction, leading to greater losses and an increased bandwidth. One can almost regard the spectral peak created by the formant pair to be a single-peaked prominence rather than a two-peaked prominence. (For an example, see Fujimura & Lindqvist, 1971.) This view leads to the speculation that rounding has the effect of creating a single-peaked spectral prominence whereas for an unrounded vowel the prominence created by proximity of F_1 and F_2 or of F_2 and F_3 has a double peak. (The perceptual consequences of prominences that result from of two closely-spaced formants are discussed in Section 5.1.2 below.) This influence of rounding, then, would be another example of an articulatory–acoustic relation that displays a shift from one type of acoustic pattern to another as an articulatory parameter passes through a particular region.

2.6. *Coupled resonators and consonants*

Proximity of two formants as a consequence of weakly coupled resonators, resulting in acoustic stability with changes in constriction position, appears to play a role in selecting articulatory configurations for consonants as well as for vowels. The cross-sectional area of the constriction for consonants is usually less than it is for vowels. When the constriction is formed with the tongue body, Fig. 3 shows that there is a constriction position about two-thirds of the distance from the glottis to the lips for which the second and third formant frequencies are relatively close together. This tendency for convergence of F_2 and F_3 appears to be utilized as one way of creating the narrow midfrequency spectral prominence that characterizes a velar consonant. The prominence resulting from this proximity in F_2 and F_3 in the region of the transition toward or away from a velar closure is especially evident when there is no opportunity to provide information concerning a midfrequency prominence in a burst. An example is a velar nasal consonant, a spectrogram of which is shown in Fig. 14. The proximity of F_2 and F_3 at both the implosion and the release of the velar in the words "sang it" is evident. Perturbations in the place of articulation for the velar consonant are not expected to modify greatly

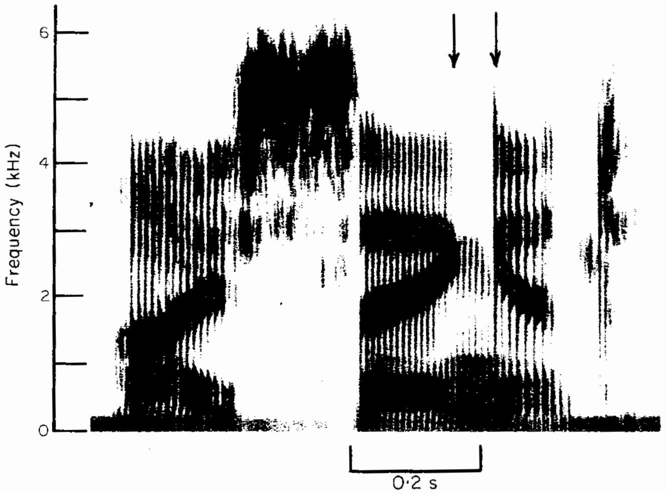

Figure 14. Spectrogram of the phrase "I sang it", illustrating the proximity of F_2 and F_3 immediately adjacent to the closure interval for /ŋ/ (marked by arrows).

the basic acoustic attribute of a midfrequency spectral prominence that is a consequence of the two converging formants.

In Figs 3 and 4 there is also a region in which the constriction is a few centimeters above the glottis, where F_1 is relatively high and F_2 is a minimum. This is evidently the region used for producing the tongue-body constriction for a pharyngeal consonant. Such a consonant is characterized by a high F_1 and a low F_2 in the transitions of the vowel adjacent to the pharyngeal constriction (Klatt & Stevens, 1969; Alwan, 1986). Examples are given in Fig. 15, which shows spectrograms of syllables in which the voiced pharyngeal consonant /ʕ/ is followed by the vowels /ɑ/ and /i/. Again the basic property of a closely spaced pair of formants is expected to be relatively insensitive to perturbations of the constriction position in this lower pharyngeal region.

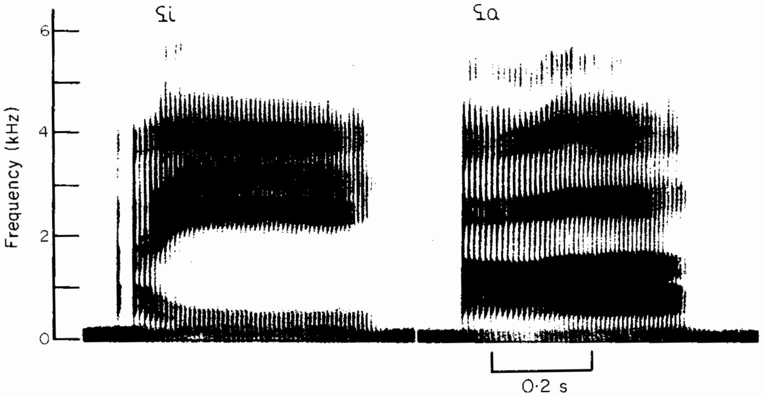

Figure 15. Spectrograms of the syllables /ʕi/ and /ʕa/ produced by a speaker of Arabic. The proximity of F_1 and F_2 in the initial pharyngeal consonant is evident.

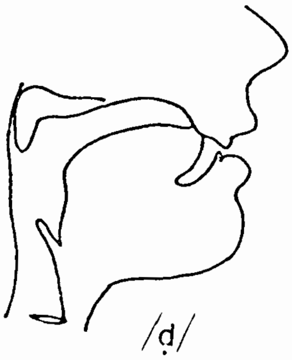

Figure 16. Midsagittal vocal-tract configuration for the retroflex consonant /ḍ/, illustrating the contribution of the space under the tongue blade to the lengthening of the front cavity.

Proximity of two higher formants as a consequence of weakly coupled resonators can also be realized by forming a constriction with the apex of the tongue close to the palate. For example, if the tongue blade is raised while the tongue tip is placed close to the palate, a space is formed under the tongue blade, as shown in Fig. 16. The dimensions of this space are such that the natural frequency of this front cavity can be adjusted to bring F_3 and F_4 close together. For the simple model of Fig. 2(a), with the corresponding formant patterns in Fig. 3, this point is located at about $l_1 = 11.2$ cm. That is, the lowest resonance of the front cavity becomes approximately equal to the third natural frequency of the back cavity, creating a prominent spectral peak in the sound output, with characteristics that are somewhat insensitive to the exact position of the constriction, at least to within 1–2 mm. The spectrogram of a retroflex stop consonant shown in Fig. 17 illustrates this coming together of F_3 and F_4.

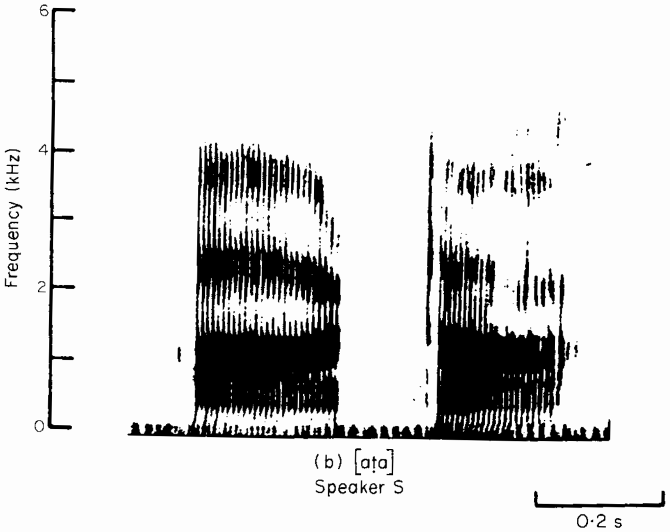

Figure 17. Spectrogram of the utterance /aṭa/ produced by a Hindi speaker. The spectrogram illustrates the proximity of F_3 and F_4 immediately preceding the consonant implosion.

With a more extreme retroflexion of the tongue blade, accompanied by some lip rounding, the natural frequency of the cavity in front of the constriction can be lowered still more. With appropriate adjustment of the shape of the back cavity, the front- and back-cavity resonances can be manipulated so that it is F_2 and F_3 that are close together rather than F_3 and F_4. The lowered frequency of the back-cavity resonance is achieved in part by forming a secondary constriction with the tongue body midway between the glottis and the constriction formed with the tongue blade. This is the configuration that is used to produce one type of /r/ in English.

These examples of different consonantal configurations suggest that there are several places along the vocal tract where a constriction can be located such that a natural frequency of the cavity downstream from the constriction is close to a natural frequency of the cavity upstream from the constriction. These constriction locations lead to well-defined tendencies in the pattern of formant frequencies observed at the implosion or at the release of the consonant, and this pattern is only weakly sensitive to perturbations in the placement of the constriction.

As the relations in Figs 3 or 4 would suggest, a shift in constriction position of 1–2 mm will produce a larger change in the formant frequencies at the $F_3 - F_4$ proximity point (e.g., at $l_1 = 11.2$ cm in Fig. 3) than at the $F_1 - F_2$ proximity point (e.g., at $l_1 = 1$ cm in Fig. 4). Presumably, however, the motor and orosensory system is capable of greater precision in positioning a constriction with the tongue blade in the region of the hard palate than with the tongue body in the lower pharynx.

When the constriction is located in the region of or anterior to the alveolar ridge, the frequency of the front cavity resonance is too high to participate with one of the back-cavity resonances in forming a perceptually significant spectral prominence. Thus the properties of alveolar and labial consonants are not determined by these considerations relating to coupled acoustic cavities. This attribute of these consonants has led to their classification as *diffuse* by Jakobson, Fant & Halle (1963), or as [+anterior] by Chomsky & Halle (1968).

3. Sounds produced with turbulence noise at a constriction

3.1. *Amplitude of turbulence noise*

When an obstruent consonant is produced, turbulence noise is generated in the rapid airflow in the vicinity of the constriction. The amplitude and spectral characteristics of this noise depend on the airflow and on the characteristics of the constriction and of any obstacles or surfaces immediately downstream from the constriction. This noise source provides an excitation for the vocal tract, and the sound radiated from the lips can be viewed as the product of the source spectrum, the transfer function of the vocal tract, and the radiation characteristic (Fant, 1960).

A midsagittal section of the vocal tract with a supraglottal constriction is shown in Fig. 18, together with a schematic representation of this configuration as a tube with a constriction and an obstacle downstream from the constriction. When a subglottal pressure P_s is applied, an airflow U passes through the vocal tract, and this airflow depends upon the cross-sectional areas A_g and A_c of the glottal opening and of the supraglottal constriction. This relation is given approximately by

$$P_s = \frac{\varrho U^2}{2A_g^2} + \frac{\varrho U^2}{2A_c^2} \qquad (4)$$

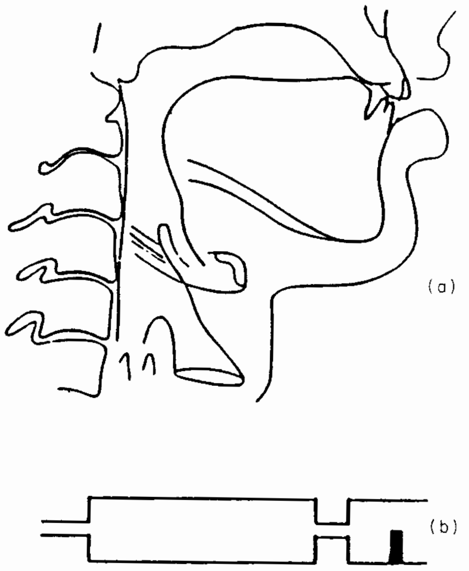

Figure 18. (a) Midsagittal configuration of vocal tract for a fricative consonant. (b) Mechanical model of fricative noise generation with an obstacle downstream from the constriction.

where ϱ = density, as long as the areas are not sufficiently small that the viscous resistance becomes significant in relation to the dynamic resistance terms given here. Experimental data and theoretical analysis show that the turbulence noise source at a constriction in the vocal tract can be modeled as a sound-pressure source p_s (Fant, 1960; Shadle, 1985). If the cross-sectional area of the constriction is A, the amplitude of this source is given approximately by

$$p_s = KU^3A^{-5/2}. \qquad (5)$$

This equation was derived in part from theories of turbulence noise generation and in part from examination of experimental data of Shadle (1985). The proportionality constant K is dependent on the shape of the constriction and on the configuration of surfaces against which the airflow impinges downstream from the constriction. In fact, the presence of an obstacle appropriately positioned in the airflow downstream from the constriction can cause an increase in the amplitude of the sound source by as much as 20 dB relative to that for the non-obstacle condition (Shadle, 1985). Such an obstacle (formed by the lower incisors, for example) is utilized to form *strident* consonants.

For a given subglottal pressure and glottal area it is possible to compute the relative amplitudes of the noise sources at the glottal and supraglottal constrictions as the cross-sectional area of the supraglottal constriction is manipulated. The calculation makes use of the above equation (4) relating P_s, A_g, A_c, and U, together with Equation (5) giving the relative amplitude of the noise source. The results of these calculations are shown in Fig. 19(a). The two curves in the figure give the relative amplitudes of the noise sources at the two constrictions as the cross-sectional area of the supraglottal constriction is increased. The feature of interest is the broad maximum in the curve for the supraglottal noise source. In this example, the amplitude of the noise source is within 3 dB of its maximum value over a range of constriction sizes from 0.03 to 0.2 cm^2. For very small

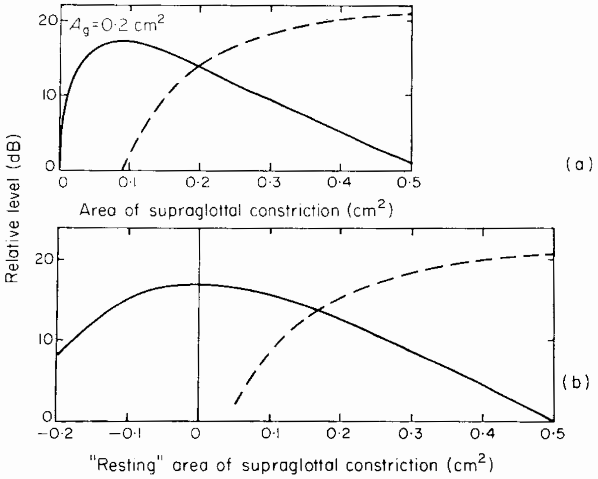

Figure 19. (a) Calculated levels of the turbulence noise sources at the supraglottal and glottal (dashed line) constrictions as a function of the area of the supraglottal constriction. The area of the glottal constriction is fixed at 0.2 cm^2. (b) Same as (a), except that the size of the constriction is modified by the presence of the intraoral pressure. The abscissa is the area that the supraglottal constriction would assume before application of the subglottal pressure (from Stevens, 1987).

constriction sizes, the source amplitude decreases because of the reduction in the airflow. The decrease for large constriction sizes occurs as the airflow becomes limited by the glottal constriction rather than by the supraglottal constriction. This is another example, then, of the relative stability of an acoustic parameter (in this case the amplitude of the turbulence noise source) with changes in an articulatory parameter (in this case the cross-sectional area of the constriction).

The analysis of the noise generated at the vocal-tract constriction in Fig. 18 can be carried further if we note that the walls of the vocal tract behind the constriction show some outward displacement in response to an increased intraoral pressure. A consequence of this property of the vocal-tract walls is that the cross-sectional area A_c of the constriction is determined not only by the passive adjustment of the supraglottal structures to produce a particular "resting" configuration of the constriction in the absence of an applied pressure, but also by forces due to the pressure behind the constriction. The detailed analysis of this situation is beyond the scope of this paper and involves calculation of the displacement of the surfaces near the constriction from knowledge of the intraoral pressure, the mechanical compliance of the walls, and the shape of the constriction. Figure 19(b) shows a revised plot of the type given in Fig. 19(a) in which the abscissa is the "resting" cross-sectional area of the constriction. The general form of the two figures is the same, but the maximum in the noise amplitude for the supraglottal constriction in Fig. 19(b) is considerably broader. For example, even when the resting cross-sectional area is set to zero, the application of pressure opens up the constriction to cause airflow and noise generation. (Negative resting areas in the figure represent constriction configurations for which the opposing structures that form the constriction exert some force on each other in the absence of an intraoral pressure.)

The relations between noise amplitude and constriction size in Fig. 19 can provide some insight into the mechanism of production of a fricative consonant in intervocalic position. As the active articulator moves from the relatively unconstricted position for the vowel to a constricted configuration for the consonant, the noise generated near the constriction increases to a maximum and its amplitude remains essentially constant as the constriction size decreases, the articulator reverses its direction of movement, and the constriction size begins to increase. The articulator can undergo a continuous smooth movement, and there is no requirement that it remain in a fixed position over the duration of the fricative consonant in order to maintain a constant amplitude of noise. As Fig. 19(b) shows, the fact that the vocal-tract walls are compliant contributes much to this stability of the noise amplitude.

The situation is somewhat different for a voiced fricative consonant, where there is a requirement that vocal-fold vibration occur at some time during the consonant interval. If vocal-fold vibration and frication noise generation are to occur simultaneously, a more precise adjustment of both the glottal configuration and the supraglottal constriction is needed in order to maintain an intraoral pressure that is intermediate between the subglottal pressure and atmospheric pressure (Stevens, 1987). Deviation from this adjustment can lead either to devoicing or to significant reduction in the amplitude of the frication noise. It frequently happens that a "voiced" fricative consonant is produced with vocal-fold vibration over only a portion of the consonantal interval, with frication noise reaching its maximum amplitude during the voiceless portion of the interval. Simultaneous generation of vocal-fold vibration and frication noise is inherently unstable, and special adjustments, often in the timing of these events, are used to signal the feature of voicing.

3.2. *Constriction position for obstruent consonants*

As the position of the constriction in the vocal tract is manipulated between the glottis and the lips, the sound radiated from the lips for an obstruent consonant is determined by the configuration of the vocal tract downstream from the constriction as well as by the characteristics of the turbulence noise source in the vicinity of the constriction. The source can usually be considered as a sound-pressure source distributed over a region of the vocal tract downstream from the constriction. Its spectrum depends to some extent on the constriction size and the airflow. However, for the situations normally encountered for fricative consonants, the spectrum is relatively smooth, and decreases above 1 kHz with a slope that gradually increases from 6 dB/octave to about 12 dB/octave at 5 kHz. An estimate of this source spectrum, as determined by Shadle (1985), is given in Fig. 20(a).

The model of Fig. 20(b) can be used as a basis for examining the broad characteristics of sound generated by turbulence noise when a constriction is positioned at various points along the vocal tract. We shall assume that the source in the model is concentrated at distance l_s downstream from the constriction, and that this distance remains the same independent of the constriction position. Furthermore, the source strength is assumed to be independent of constriction location. These are only rough approximations to the situation in the real vocal tract for obstruent consonants, and we will refine this initial analysis later. However, the model will indicate the gross features of the output as the constriction position is manipulated. More precise adjustments of the model to provide a better simulation of vocal-tract behavior will alter the details but not the gross attributes of the radiated sound.

The transfer function from the source sound pressure p_s to the volume velocity U at the mouth has poles at the frequencies of the front-cavity resonances, and has a zero or

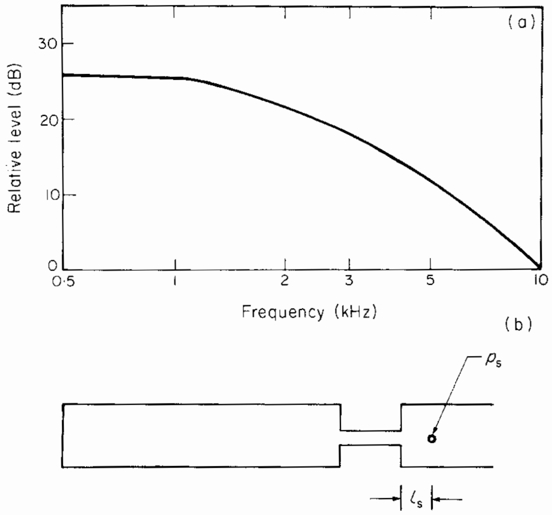

Figure 20. (a) Spectrum of sound-pressure source for a configuration similar to that in Fig. 18(b) for an airflow of about 400 cm^3/s. Diameter of (circular) constriction is 0.32 cm, and distance from constriction to obstacle is 3 cm (after Shadle, 1985). (b) Model used for examining characteristics of sound generated by turbulence noise for various positions of the constriction. Distance from constriction to location of source p_s is indicated by l_s.

near zero frequency. We assume that l_s is small compared with a wavelength in the frequency range of interest, such that zeros at higher frequencies do not influence the transfer function. The calculated spectrum of the radiated sound pressure for several constriction positions with l_s = 1 cm, is shown in Fig. 21. The main feature of interest is the fact that, for constriction positions within a few centimeters of the mouth opening, there is a maximum in the amplitude of the spectral prominence corresponding to the front-cavity resonance. When the front cavity is shorter, the amplitude of the spectral prominence decreases primarily because of the increased losses arising from radiation at high frequencies. For longer front cavities, the decreased amplitude is due to the decreased magnitude of the transfer function as the natural frequency decreases, particularly since there is a smaller degree of coupling of the source to the front-cavity resonance. In terms of maximizing the sound output, then, there is an optimum position for the constriction. This position corresponds roughly to the constriction formed by the tongue blade for the consonant /s/ or /t/.

The overall pattern in Fig. 21 is modified in several ways if the model is adjusted to represent the vocal tract in greater detail. When the constriction is located a few millimeters posterior to the incisors, the lower teeth form an obstacle against which the jet of air from the constriction impinges. The amplitude of the noise source is enhanced by the increased turbulence at the lower incisors. Thus in the region of Fig. 21 where the prominence is greatest, the amplitude of the spectral peak is even greater than that shown in the figure, and the maximum is enhanced. In fact, experimental data and theoretical analysis show that in this region the high-frequency spectrum amplitude of the radiated sound exceeds the spectrum amplitude in the same frequency range for a vowel produced with about the same subglottal pressure. This enhanced high-frequency spectrum amplitude in relation to the vowel for obstruents is presumably the basis for the distinction

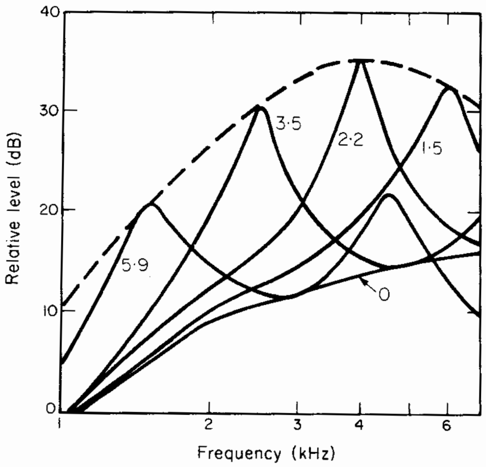

Figure 21. Spectrum of radiated sound pressure for a fricative consonant modeled by the configuration in Fig. 20(b), and with a source spectrum as in Fig. 20(a). Different curves correspond to different constriction positions or front-cavity lengths as indicated (in cm). The source is assumed to be located at an obstacle 1 cm downstream from the end of the constriction. The front-cavity cross-sectional area is 1 cm^2, and calculated radiation losses are based on this radiating area. Losses at the constriction are based on a constriction of 0.2 cm^2 and a subglottal pressure of 8 cm H_2O. The dashed line gives the envelope of the spectrum amplitude at the peak.

between coronal consonants (or non-grave consonants, using the terminology of Jakobson *et al.* 1963) and non-coronal (or grave) consonants.

Another modification to Fig. 21 occurs if the influence of the back cavity is taken into account. As we have seen in Fig. 3, the natural frequencies of the two parts of the tube are shifted somewhat when they are coupled through a constriction, and the amount of shift depends on the size of the constriction. At a constriction position for which the uncoupled front and back cavity resonances are equal (such as $l_1 = 9.3$ cm in Fig. 3), the two natural frequencies for the coupled resonators cannot be assigned to one cavity or the other, but the energy for each mode is stored about equally in the two parts of the system. In this ideal case, then, both resonances are excited by a source located near the designated point in Fig. 3. The lower of the two frequencies shows a more prominent spectral peak, however, since it corresponds to the symmetric mode, which has less acoustic loss and hence a narrower bandwidth. For constriction positions to the left or to the right of this crossover point it is the natural frequency close to the uncoupled front-cavity resonance that receives greatest excitation by the noise source. The situation is schematized in Fig. 22, which shows the resonance that is most strongly excited by the noise source as the constriction position is manipulated. Near the intersection points there is a rather abrupt jump in the frequency that receives greatest excitation by the source. Between the intersection points, the spectral peak of the radiated sound corresponds to a particular formant number. As the constriction position is displaced forward in the vocal tract, then, the prominences in the spectrum of the sound output when there is turbulence noise at the constriction undergo rather abrupt modifications, as the resonance of the front cavity changes from being associated with F_2 to F_3 to F_4 and above.

A special case of this rather abrupt change in the formant number with which the front cavity is affiliated can be observed when the position and configuration of the tongue

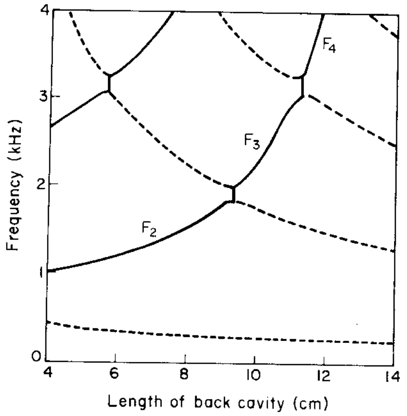

Figure 22. Natural frequencies of a configuration like that in Fig. 2(a), with $A_1 = A_2 = 3\,\mathrm{cm}^2$ and $A_c = 0.2\,\mathrm{cm}^2$, as a function of l_1. Total length of tube is 16 cm. The solid line indicates the frequency of the front-cavity resonance, which is the natural frequency that is most strongly excited by a noise source downstream from the constriction. At the points where front- and back-cavity resonances are close together, there is a jump in the formant number associated with the front-cavity resonances, as shown by the discontinuities in the solid lines.

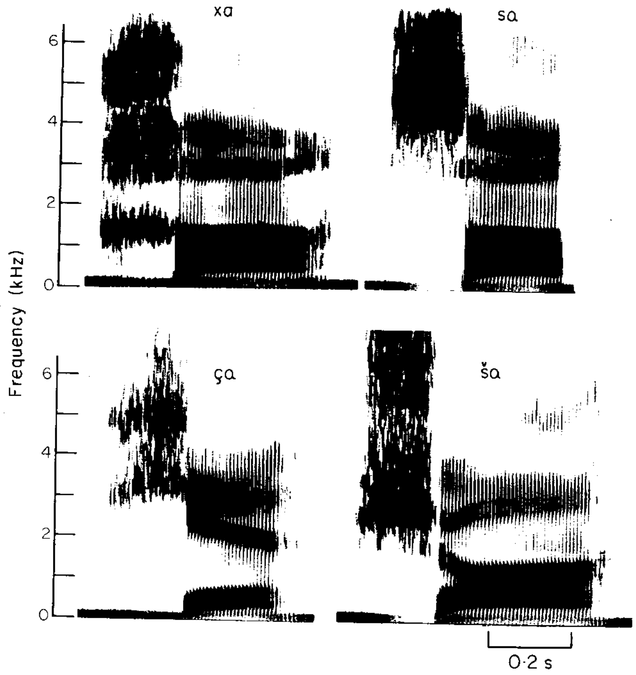

Figure 23. Spectrograms of fricatives produced with constrictions at several points along the vocal tract: near the soft palate (upper left), in the region of the hard palate (lower left), near the alveolar ridge (upper right), and in the palatal region with a space under the tongue blade (lower right). The lowest formant affiliated with the front cavity shifts from F_2 to F_3 to F_4 or F_5 as the constriction is moved to a more anterior position. Phonetic symbols corresponding to the different consonants are indicated on the spectrograms.

blade is shifted from /s/ to /š/. As the tongue blade is displaced to a more posterior position, it appears that, for many speakers, a space is created between the lower surface of the tongue blade and the floor of the mouth (Perkell, Boyce, & Stevens, 1979). Creation of this space results in the rather abrupt introduction of a longer cavity anterior to the constriction. This cavity is sufficiently long that it usually becomes affiliated with the third natural frequency of the vocal tract, rather than the fourth or fifth natural frequency that is characteristic of /s/. The perceptual consequence of this introduction of a spectral prominence for the frication noise in the F_3 region has been confirmed (Stevens, 1985), as discussed in Section 5.3 below.

When an obstruent consonant is adjacent to a vowel, the prominences in the spectrum of the consonant show continuity with particular formants in the vowel, and it is possible to associate the prominences with those formants. The changes in cavity affiliations of the formants for different places of articulation for fricative consonants are illustrated in Fig. 23. For the consonant /χ/, the lowest spectral prominence is associated with F_2, whereas F_3 becomes the lowest front-cavity resonance for /ç/. In the case of /s/ the front cavity is much shorter, and is associated with F_4 or F_5 of the adjacent vowel. The excitation of F_3 for /š/ is evident in the spectrogram of /ša/.

4. Vocal-fold vibration

When the vocal folds are positioned close together and a pressure is applied across the glottis, the folds are set into vibration. This mechanical vibration modulates the flow of air through the glottis, and this modulated airflow forms acoustic excitation of the vocal tract. The initiation and maintenance of vibration requires that the vocal-fold surfaces have a particular range of values of stiffness, the degree of adduction or abduction of the glottis is within a certain range, and there is a sufficiently large transglottal pressure. When these parameters are outside of these ranges, vocal-fold vibration will cease. Thus, for example, if the glottal width roughly exceeds 2-3 mm, or the transglottal pressure drops below 2-3 cm H_2O (Ladefoged, 1967), the vocal folds will tend not to vibrate.

Evidently, then, vocal-fold vibration occurs over a range of values of the aerodynamic and mechanical parameters, and there is no vibration when the parameters are outside this range. There may be a narrow region where the vocal folds vibrate weakly, or where vibration cannot be initiated, but, once started, can be maintained. Basically, however, the vocal folds with their accompanying aerodynamic forces can be in two states: vibration or no vibration. In the case of obstruent consonants, the time interval over which the presence or absence of voicing is manifested is usually the entire obstruent region, although vocal-fold vibration is sometimes limited to the few tens of milliseconds adjacent to the boundary with a sonorant region.

When the vocal folds are in a condition that leads to vibration, the manner in which they vibrate and the waveform of the resulting airflow modulations can be manipulated through finer adjustments of the glottal configuration and vocal-fold state (Ladefoged, 1983). We still lack a sufficiently detailed understanding of the mechanism of vibration and its acoustic consequences to be able to delineate quantitatively the kinds of glottal adjustments that will lead to significant changes in the properties of the sound output—changes that listeners are capable of utilizing to make distinctions in language. Whether there are distinct or qualitatively different "modes" of vibration, and the nature of these modes, if they exist, is at present a matter of speculation.

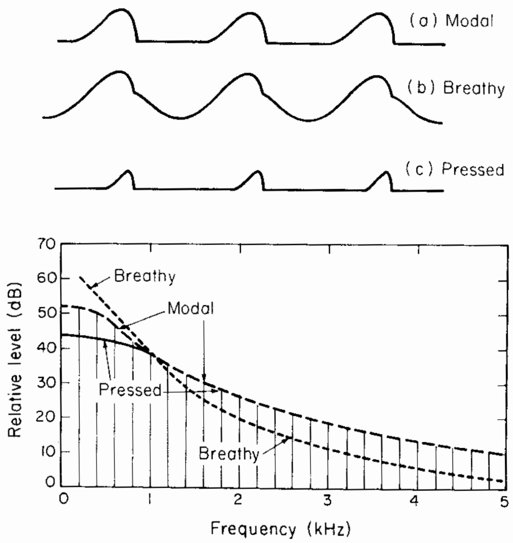

Figure 24. The upper part of this figure shows schematized waveforms of glottal airflow for three types of vocal-fold vibration as indicated. The lower panel displays spectra for these three waveforms. Harmonics are shown for a 200 Hz fundamental frequency, and the spectrum envelope for each type of phonation is indicated.

One such speculation is that there is a basic "modal" type of vibration, and that this modal vibration can be modified to yield a "breathy" and a "pressed" phonation through appropriate abducting or adducting maneuvers of the folds. (One attempt to quantify these aspects of vocal-fold vibration is given in Stevens, 1988.) In the modal state the vocal folds are positioned so that they are touching lightly along their entire length in the absence of an applied subglottal pressure. When the folds are vibrating in this state, the waveform of the vibration is determined in part by the mass and stiffness of the folds. Glottal closure during the cycle of vibration tends to occur along the entire length of the membranous folds, and there is an abrupt discontinuity in the airflow waveform at the instant of closure. The waveform is shown schematically in Fig. 24(a), and a sketch of the spectrum envelope of the waveform is given in the lower panel as a dashed line.

For the breathy type of phonation, the vocal folds are abducted along a portion of their length near the arytenoid cartilages. The anterior portions of the folds vibrate in a manner similar to modal vibration, whereas the motion in the posterior portions is such that the vocal folds do not touch during the cycle. Consequently the amplitude of motion may be greater, and there is no discontinuity in the waveform of the airflow that passes through this posterior portion of the glottis. This component of the waveform is approximately sinusoidal, whereas the component contributed by the anterior portion has a waveform similar to that for modal vibration. The amplitude of the part of the glottal pulse that is contributed by the anterior portion is somewhat less than the amplitude of the pulse for modal vibration, and consequently the high-frequency spectrum amplitude for breathy phonation is less than it is for modal vibration. The waveform for this condition is schematized in Fig. 24(b), and the spectrum envelope (dotted line) in the bottom panel shows an increased amplitude at low frequencies and a reduced amplitude at high frequencies. The increased amplitude of the fundamental component in the spectrum (Bickley, 1982; Ladefoged, 1983) is a consequence of the tendency for a sinusoidal waveform of vibration in the abducted posterior portion of the glottis where the vocal folds do not achieve closure during the cycle.

In the case of the pressed type of phonation, the vocal folds are tightly adducted. It is reasonable to assume that the tissue remains under compression during the cycle of vibration for this type of phonation. The mechanical movements of the vocal-fold tissue within a cycle consist more of a change in shape of the cover of the vocal fold or a redistribution of the tissue rather than a lateral movement of the entire fold. The displacement of the tissue under this condition is determined largely by the internal resistance of the tissue rather than by its mass. As indicated in Fig. 24(c), the glottal pulses for pressed phonation are narrow, the peak airflow is reduced relative to that for modal phonation, and there is a relatively long closed phase. The spectrum envelope is below that for modal vibration at low frequencies, and is shown in the lower panel as a solid line. At higher frequencies the spectrum envelope is about the same for modal and pressed vibration.

If there is any substance to these speculations, then we would have another example of an acoustic–articulatory relation characterized by regions in which the acoustic parameter undergoes significant changes when an articulatory parameter (in this case an adduction–abduction maneuver) passes through certain critical regions. These critical regions occur (1) when the degree of abduction is such tht the posterior portions of the vocal folds do not touch during the cycle, (2) when the vocal folds are just touching along their length and the motion is determined primarily by the mass and stiffness of the tissue, and (3) when the degree of adduction is such that the vibration is in a mode where the vocal folds remain under compression, and the pattern of mechanical movement consists of a redistribution or change in shape of the vocal folds, with little net lateral movement of the center of mass.

5. Quantal effects in auditory processing

In our discussion to this point, we have given a number of examples showing that manipulation of a particular articulatory parameter through a succession of values gives rise to a sound output in which certain attributes of the sound do not change in a monotonic fashion. Rather, there tend to be ranges of the articulatory parameter for which there is a substantial shift in the acoustic attribute, and other ranges in which the acoustic property is relatively insensitive to perturbations in the articulatory parameter. Ultimately, however, we are interested not in the physical acoustic parameter of the sound but in some aspect of the response of a listener to the acoustic parameter. The existence of relations of the type shown in Fig. 1 between some parameter of the auditory response and an articulatory parameter can be a consequence of either a quantal acoustic–articulatory relation or an auditory–acoustic relation, or possibly a simultaneous occurrence of quantal relations in both domains.

We turn now to examine the results of some experiments that assess how the auditory or psychoacoustic responses of listeners (human or animal) to particular stimuli change when the acoustic parameters that describe the stimuli are manipulated. Is there evidence for lack of monotonicity in the relations between listener responses and acoustic parameters of the type that occur in speech, such that for changes in the acoustic parameters over one part of the range there are large changes in response, whereas over other parts of the range these changes are small? The evidence in this auditory domain tends to be less direct and hence less compelling than evidence from the articulatory–acoustic domain, because less is known about auditory responses to speechlike signals.

Evidence for auditory–acoustic relations can be of several kinds. One type of data comes from studies of the response of the auditory nerve to speechlike sounds. The representation of sounds at the level of the auditory nerve is a transformation of the acoustic signal, and it is of interest to examine whether this transformation tends to make the pattern of auditory-nerve responses for certain sounds more similar and to make the pattern for other sounds more distinct or well separated.

Other potential evidence comes from psychophysical data in which identification, discrimination, or similarity judgements are obtained from listeners as particular attributes of speechlike stimuli are manipulated. For example, a number of experiments have found that there is lack of monotonicity in discrimination of small changes in some acoustic parameter as that parameter takes on different values along a continuum (see, for example, several papers on categorical perception of speech in Harnad, 1987). This evidence is derived from experiments with normally hearing adult listeners, from studies with infants (who presumably are not strongly influenced by exposure to phonetic contrasts in language), and from research with animals. We shall not attempt here to review these psychophysical data and their interpretation, and will give only a few examples.

5.1. *Vowel perception*

5.1.1. *Spectral prominences*

As is well known, one of the attributes that distinguish vowels from consonants is that the spectra of vowels are characterized by several prominent peaks, particularly in the midfrequency range of about 800–3000 Hz. (A spectral peak corresponding to F_1 can also occur at a frequency below this range for vowels, but under some circumstances the degree of prominence is reduced by widening and flattening the peak.) The spectra that are observed during the production of consonants tend to be more diffuse, such that any spectral peaks that occur are less prominent within this frequency range, or the spectral peaks are outside of this frequency range. There are, however, exceptions to this general statement.

Data from studies of the responses of auditory-nerve fibers when the stimulus is a pure tone or is a vowel-like sound with prominent spectral peaks show that the responses of all of the fibers with characteristic frequencies in the vicinity of a spectral prominence in the stimulus exhibit synchrony of firing to the frequency of the prominence rather than to their own characteristic frequencies (Kiang, Watanabe, Thomas & Clark, 1965; Johnson, 1980; Young & Sachs, 1979; Delgutte & Kiang, 1984*a*). This result is not unexpected, since the frequency of the mechanical response at a point on the cochlear partition when the stimulus has a narrow spectral prominence will be dominated by the stimulus frequency rather than by the characteristic frequency associated with that point. This property stems from a basic attribute of the response of simple linear bandpass filters.

When the stimulus is a brief click, the spikes in the responses of auditory-nerve fibers with characteristic frequencies in the range up to about 3 kHz tend to show firings that are synchronous to their own characteristic frequency (Kiang *et al.*, 1965). Again this type of response is not unexpected, since the portion of the cochlear partition with which a given fiber is associated shows a mechanical response that is dominated by its characteristic frequency. The fiber will tend to fire when there are peaks in one direction in the mechanical response, and hence will show synchrony to this frequency. It is expected also that the responses of fibers with about the same characteristic frequency when the stimulus is broad-band noise will show a predominance of time intervals equal to the reciprocal of the characteristic frequency.

Depending on the spectral characteristics of the stimulus, therefore, we observe two distinctively different patterns of responses of auditory-nerve fibers: synchrony of a fiber to its own characteristic frequency or synchrony of groups of fibers to the frequency of a spectral prominence in the stimulus. We can imagine how this pattern of response would change as the characteristics of the stimulus are modified, starting with a sound with a narrow spectral prominence and gradually increasing the width of this prominence until its spectrum becomes relatively flat. As long as the bandwidth of the prominence is less than the effective bandwidth of the peripheral analyzing filters, the vibration of the cochlear partition in the region that responds maximally to the stimulus frequency will be "captured" by the stimulus frequency. The synchrony of firing of auditory-nerve fibers associated with this region will be dominated by this frequency. When the bandwidth of the stimulus exceeds that of the peripheral analyzing filters, the stimulus frequency no longer dominates the response of the cochlear partition, and the auditory-nerve fibers revert to firing synchronously at their own characteristic frequencies. Data from auditory physiology show that this synchronously-responding property of auditory-nerve fibers is strongly evident for characteristic frequencies up to 2–3 kHz, and the ability to respond synchronously becomes gradually attenuated at frequencies above 3 kHz (Johnson, 1980).

This behavior of the peripheral auditory system provides, then, an example of an auditory–acoustic relation that shows a "threshold" effect, or a relatively abrupt shift in the response pattern as certain parameters of the stimulus pass through a critical range. Whether at high levels in the auditory system the human listener has a mechanism for focusing on this contrasting response pattern is not known. The existence of this pattern at the level of auditory nerve, however, would suggest that this contrast between the two types of response would be utilized at higher levels.

The bandwidths of the formants for vowels are usually less than the bandwidths of the auditory filters in the frequency range up to about 3 kHz—the range where bandwidth data are available. Figure 25 gives the commonly accepted bandwidths of the auditory filters (Zwicker, 1961), together with estimates of average formant bandwidths as a function of frequency (Fant, 1972). The formant bandwidths are considerably less than the bandwidths of the auditory filters, particularly in the frequency range 800 to 3000 Hz. At lower frequencies, the formant bandwidths also fall below the critical bandwidths for most normally voiced non-nasal vowels. Wider effective first-formant bandwidths can be achieved, however, for nasal vowels, for breathy-voiced vowels, and possibly for vowels produced with an advanced tongue root.

5.1.2. *Proximity of two formants for vowels*

As the frequencies of the first two or three formants for a vowel are manipulated, the vowel quality changes, and listeners identify different ranges of these formant frequencies with different vowels, depending on their language. Chistovich and her colleagues (e.g., Chistovich, Sheikin & Lublinskaya, 1979; Chistovich & Lublinskaya, 1979) have examined

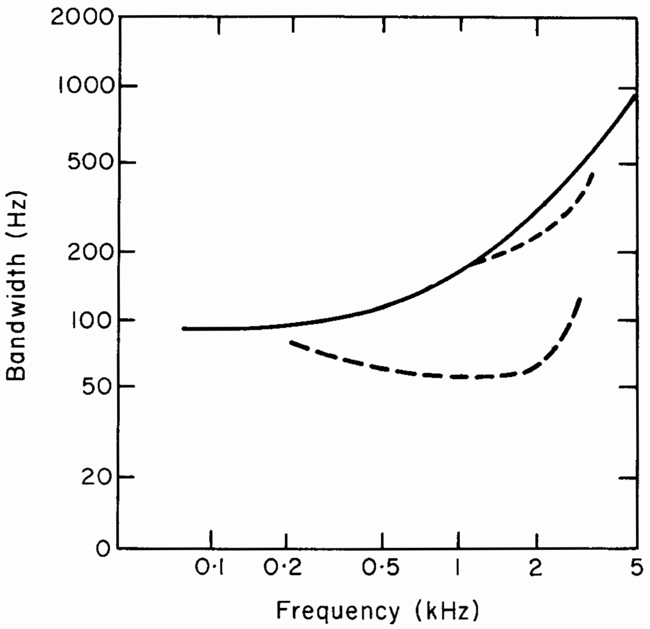

Figure 25. The solid line gives the width of the critical band as determined from psychophysical experiments (after Zwicker, 1961). The dashed line shows ranges of estimates of the bandwidth of vowel formants as a function of frequency, for non-nasal vowels with modal voicing. The short-dashed line gives estimates of the bandwidth of a noise burst that is generated with a constriction formed by the dorsum of the tongue at different points against the hard or soft palate, leading to different natural frequencies of the front cavity.

some aspects of the response of listeners to synthetic two-formant vowels, and have found that there is a change in the response when the frequencies of the two formants are sufficiently close.

In the basic experiments, a number of two-formant vowels were synthesized with different spacings between the formants and with different relative amplitudes of the formants. Listeners were asked to adjust the frequency of a one-formant vowel-like sound to be as close as possible in quality to each of the various two-formant vowels. The results showed that when the spacing between the two formants was greater than a certain critical value, the listeners adjusted the frequency of the matching stimulus to be equal to one or other of the two formants in the experimental stimulus, or else there was large variability in matching performance. When the spacing was less than this critical value, the frequency of the formant in the matching stimulus was at some point between the two formants. The location of the matching formant depended on the relative amplitudes of the two spectral peaks in the experimental stimulus. The critical spacing that represented the boundary between the two types of behavior was about 3.5 Bark.

In their reports of matching experiments with two-formant vowels, Chistovich and her colleagues also noted that, as the relative amplitude of one of the peaks is reduced when the formant spacing is less than the critical separation, the responses of a listener to the stimulus pass through two steps. The first, as noted above, is that the frequency of a one-formant matching stimulus shifts from a value intermediate between the two formants to a value equal to the frequency of the higher amplitude peak. At this stage, however, the weaker formant continues to play a role in determining some aspects of the vowel quality. The second step is that the weaker peak passes through a threshold such that it is no longer detectable. That is, a listener cannot distinguish between the one-formant stimulus and the two-formant stimulus in which one of the peaks has a low amplitude.

During the production of rounded vowels and some sonorant consonants, two formants come very close together, and this proximity is usually accompanied by a widening and hence an amplitude reduction of one of the two formant peaks, as discussed in Section 2.5 above. Examples are the proximity of F_2 and F_3 for English /r/ or for the rounded front vowel /y/. It is not unreasonable to suggest, then, that these segments are characterized by a spectral prominence that is interpreted by the auditory system as a single peak rather than a two-peaked prominence.

The results of the experiments of Chistovich and her colleagues can be viewed as another example of an auditory-acoustic relation where there are shifts in behavior (in this case a matching of two stimuli for similarity in vowel quality, or detection of the presence of a spectral peak) at certain critical ranges of values of an acoustic parameter (in this case the distance in Bark between the two formants).

The concept of a critical spacing between formants has some relevance to providing an auditory basis for the distinction between front and back vowels. It is tempting to think of back vowels as having F_2 sufficiently close to F_1 as to be within the critical spacing. For front vowels, one might expect F_2 to be high and sufficiently close to F_3 (or possibly F_3 to F_4) to be again within the critical spacing. These issues have been examined by Carlson, Granström & Fant (1970), Carlson, Fant & Granström (1975) and by Syrdal (1985) and Syrdal & Gopal (1986).

5.1.3. *Perceptual or auditory responses relating to the F_1 region for vowels*

The concept of a critical formant spacing discussed above is relevant to an understanding of the perceptual constraints underlying vowel distinctions that are signaled by proximity of F_2 to F_1 or to F_3. We turn now to a consideration of perceptual or auditory phenomena that are potential factors in the auditory interpretation of vowels in the F_1 region. Three kinds of manipulations of the spectrum in the F_1 region can be carried out by a speaker: (1) the fundamental frequency F_0 can be manipulated; (2) the first-formant frequency F_1 can be changed; and (3) the shape of the spectrum can be modified independently of the frequency of F_1. In the last case, the modification of the low-frequency spectrum can be achieved in several ways: (a) by introducing acoustic coupling to the nasal cavity through the velopharyngeal port, (b) by adjusting the laryngeal configuration, or (c) by increasing the bandwidth of F_1 by enlarging the surface area in the pharyngeal region (so as to increase the acoustic losses at the pharyngeal walls) or by creating a narrow constriction in the uvular or pharyngeal region, as for a voiced uvular consonant (Alwan, 1986).

In the case of manipulation of F_0 and F_1, several experiments have shown that the perception of vowel height is influenced by both F_0 and F_1 (for example, Traunmüller, 1981; Di Benedetto, 1987). The perceived height of a vowel is increased if F_1 is decreased and if F_0 is increased, i.e., if the difference between F_1 and F_0 is decreased. Experiments on the perception of vowels in several languages (Stevens, Liberman, Studdert-Kennedy & Öhman, 1969; Traunmüller, 1981; Di Benedetto, 1987), as well as data from acoustic analysis of vowels (Syrdal & Gopal, 1986; Fant, 1973), suggest that there is a boundary between vowel categories at a value of $F_1 - F_0$ that is equal to about 3.0 to 3.2 Bark.

For English, this is the boundary that separates the vowels /iɪʊu/ from the non-high vowels. One might speculate, then, that there is a natural perceptual boundary for this spacing between F_1 and F_0, similar to the apparent perceptual boundary related to the spacing between a pair of formants. Perceptual and acoustic data from speakers of a greater variety of languages are needed, however, in order to determine whether or not this speculation is valid, and to determine what other factors may contribute to the perception of vowel openness or height.

We consider next the perceptual effect of modifying the shape of the spectral prominence in the frequency region normally occupied by the first formant. One way in which the shape of this prominence in a vowel spectrum can be modified is through introduction of acoustic coupling to the nasal cavity. The principal effects of this coupling on the spectrum are to introduce an additional spectral peak or perturbation in the region of the first formant, to shift somewhat the frequency of F_1, and to decrease the amplitude of the F_1 peak in the spectrum (Fant, 1960). These types of perturbations can be introduced into a synthetic vowel by adding an appropriately located pole–zero pair to the vocal-tract transfer function. The amount of perturbation can be manipulated by varying the spacing between the pole and the zero (Stevens, Fant & Hawkins, 1987).

Hawkins & Stevens (1985) and Stevens, Andrade & Viana (1987) carried out experiments in which listeners from different language backgrounds made nasal–non-nasal judgements for synthetic vowels with various amounts of this pole–zero spacing. Listeners made similar identification responses independent of whether or not there was a nasal–non-nasal distinction for vowels in their language, suggesting that their responses were influenced by some basic auditory capability rather than by linguistic experience. The 50% crossover points in the identification functions occurred when the maximum perturbation introduced into the basic non-nasal F_1 prominence in the vowel spectrum was in the range 6–9 dB. This perturbation was in the form of the addition of an extra peak near F_1, a decrease in the amplitude of the F_1 peak, or both. The perturbation in the spectrum was above F_1 or was centered on F_1 for non-low vowels, and was below F_1 for the low vowel /ɑ/. Examples of spectrum envelopes in the low-frequency region for a non-nasal vowel /a/ and for a vowel that was judged to be nasal at the 50% level are shown in Fig. 26. In this case, the extra peak is below F_1, and lies about 6 dB above the canonical non-nasal spectrum.

Another type of perturbation that a talker can impose on a vowel spectrum in the low-frequency region is to modify the amplitude of the first harmonic in relation to the amplitude of the first-formant peak. In particular, adjustment of the vocal folds to a more abducted glottal configuration can lead to an enhanced amplitude of the first harmonic, as well as an increased bandwidth of the first formant. Bickley (1982) performed a series of experiments in which vowel stimuli (in a consonant–vowel frame and with a fundamental frequency corresponding to a male voice) were judged to be breathy or non-breathy by speakers of a language for which this feature was distinctive. The synthetic vowel stimuli contained various amplitudes for the first harmonic in relation to that for vowels produced with modal vocal-fold vibration. The increase in first-harmonic amplitude at the 50% crossover point in the identification functions was in the range 5–8 dB, depending on the vowel. Greater amplitudes of the first harmonic elicited judgements of the vowels as breathy. The generality of this result is temperated somewhat by the recent finding of Klatt (1987) that when the fundamental frequency is in the range appropriate for a female voice, an increased amplitude of the first harmonic tends to create a perceptual impression of nasalization rather than breathiness. In this case,

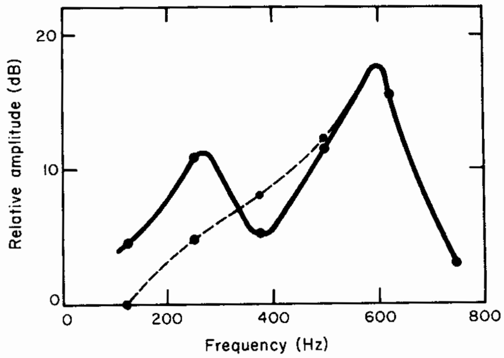

Figure 26. The dashed line shows the spectrum envelope in the low-frequency region (below 800 Hz) for a synthetic non-nasal vowel /a/. The solid line is the spectrum envelope for a vowel at the 50% point on an identification function for a non-nasal–nasal continuum from /a/ to /ã/. Nasalization was realized by introducting a pole–zero pair in the synthesizer transfer function. The points indicate the frequencies of harmonics for a fundamental frequency of 125 Hz. The maximum deviation of the solid line from the dashed curve is about 6 dB. (From Stevens *et al.*, 1987).

the higher frequency of the first harmonic approaches the range where a perturbation of the F_1 prominence is heard as nasalization.

We observe, then, some agreement in the listener responses for these two ways of perturbing the vowel spectrum at low frequencies. Modification of a prototypical modal, non-nasal spectrum in the F_1 region by 5–8 dB at the first harmonic or by a maximum of 6–9 dB for harmonics in the general region of the F_1 peak leads to a shift in listener responses to a different phonetic category. We interpret this result as another example of a threshold effect in an auditory–acoustic relation. It may be significant also that the introduction of at least some of these modifications in the F_1 prominence increases the effective bandwidth of this prominence so that it exceeds the critical auditory bandwidth, as shown in Fig. 25. More research is clearly needed, however, to determine whether these kinds of modifications in the low frequency spectrum shape can be shown to exhibit threshold effects based on some independent perceptual measure other than a simple identification test.

5.2. *Sonorancy and continuancy*

When a consonant is produced in such a way that the airway between the glottis and the output of the vocal tract has no narrow constriction, air can flow through this airway without appreciable build up of pressure above the glottis. In particular, the vocal folds can continue to vibrate in a normal fashion under these conditions, and there is no modification of the vocal-fold vibration pattern relative to that in an adjacent vowel. A consonant produced in this manner is classified as a *sonorant* consonant. Since the vocal-tract transfer function from input to output volume velocity is close to unity (or 0 dB) or slightly greater at low frequencies (below the frequency of the first resonance of the system), then the amplitude of the first and possibly the second harmonic shows very little change as the vocal tract executes a maneuver between a vowel and a sonorant consonant.

Non-sonorant consonants, on the other hand, are produced with a sufficiently narrow constriction in the supraglottal airways that the airflow from the glottis causes a pressure drop across the constriction and a consequent increase in supraglottal pressure. The resulting decrease in transglottal pressure causes a decrease in the amplitude of the glottal pulses, and, if the constriction is sufficiently narrow, the transglottal pressure decreases to the point where glottal vibration ceases. When a non-sonorant consonant is produced preceding or following a vowel, then, there is a decrease in amplitude of the glottal pulses, which is reflected in a reduced spectrum amplitude in the vicinity of the first harmonic. For example, if the intraoral pressure increases to about one-half of the subglottal pressure, a reduction of about 10 dB is expected in the amplitude of the glottal pulses, assuming that there is no active abducting or adducting laryngeal maneuver.

These considerations suggest that there is continuity in the low-frequency amplitude between a vowel and a sonorant consonant such as a nasal consonant, whereas there is a decrease in this amplitude as the closure in the vocal tract is formed for a voiced stop consonant. Measurements of a number of utterances of these types have verified this pattern of low-frequency amplitude change. The perceptual relevance of this pattern has been examined in an experiment by Blumstein & Stevens (unpublished). Consonant-vowel stimuli were synthesized in which a voiced sound with a low-frequency resonance, simulating a nasal murmur or a voice bar, preceded a vowel with transitions appropriate for an alveolar or a labial consonant. The amplitude of the low-frequency spectral peak in the consonant portion of the stimulus was manipulated through various values that were both greater than and less than the low frequency spectrum amplitude in the adjacent vowel. Listeners identified the consonants as one of /b d m n/. The crossover point in the identification function indicating nasal–non-nasal responses occurred when the low-frequency spectrum amplitude relative to that in the vowel was -1 dB for the alveolar continuum and $+2$ dB for the labial continuum. Evidently the boundary between the sonorant and the non-sonorant responses is about where the low-frequency amplitude in the consonant region is equal to that in the adjacent vowel.

Apparently, then, listeners make use of the simple property of low-frequency amplitude decreasing or amplitude not decreasing as a way of distinguishing non-sonorant from sonorant consonants, at least in the absence of additional properties (such as strong nasalization in the early part of the vowel in a syllable consisting of a nasal consonant followed by a vowel) that would provide cues that conflict with the low-frequency property. Again we have an example of a threshold effect in an auditory–acoustic relation: the threshold as exceeded when the low-frequency amplitude no longer shows an increase at the time the consonant is released into the adjacent vowel. As indicated above, a decrease in low-frequency amplitude is expected in the sound for a non-sonorant consonant, i.e., for a consonant that is produced with an increase in supraglottal pressure.

One class of consonants is produced with a complete closure at some point along the vocal tract. When this closure is released for these *non-continuant* consonants, there is an abrupt increase in amplitude in some regions of the spectrum in the mid and high frequencies. In the case where there is a build-up of pressure behind the constriction during the interval of closure, a brief acoustic transient occurs at the instant of release, produced by the initial rapid flow of air. This transient is usually followed by a burst of noise, caused by further airflow through the constriction.

An example of a phonetic contrast between a *continuant* and a *non-continuant* consonant is the fricative–affricate pair [š–č]. Acoustic analysis of the affricate [č] indicates the initial transient that occurs within the initial few milliseconds, followed by a brief drop in amplitude before the amplitude of the frication noise builds up. The fricative [š] has similar frication noise (usually somewhat longer than that for [č]) but with no initial transient.

Perception experiments that examine listener response to various manipulations of these kinds of stimuli have been carried out (see, for example, Rosen & Howell, 1987). Experiments with perhaps the closest approximation to stimuli with these characteristics have been reported by Hary & Massaro (1982), who examined identification and discrimination functions for tones with various rise times. Their stimulus continuum included tones with an abrupt initial amplitude increase characterized by an "overshoot" in amplitude, such that there was an interval following this onset in which the amplitude decreased to a steady level. This decrease occurred at various rates. The results indicated that all stimuli with an abrupt onset, including those with an initial overshoot or transient, were placed in one category (identified as "pluck") by the listeners, and those with a rise time of 15 ms or more were placed in another ("bow"). Stimuli with the abrupt onset or initial transient were well discriminated from those with a rise time of 15 ms or more, whereas stimuli within classes (different rise times, or different fall times following the initial abrupt onset) were poorly discriminated.

Apparently, then, stimuli with these different onsets are placed in different classes by listeners. Some support for this conclusion is seen in data from auditory physiology. For stimuli with an abrupt onset, particularly those with an initial transient like [č] as described above, the pattern of response of auditory-nerve fibers in appropriate frequency ranges are expected to exhibit an "overshoot" at the onset, partly as a consequence of adaptation (Kiang *et al.*, 1965). Stimuli with a less abrupt onset (such as [š]) would show little or no overshoot in the auditory-nerve response (Delgutte & Kiang, 1984*b*).

We still have much to learn about the way listeners respond to abrupt onsets and offsets, such as the role of the duration of the lower-amplitude time interval between an offset and an onset or the duration of the noise interval following the affricate release. Further experiments using stimuli with appropriate onset transients and various temporal characteristics are needed.

In summary, then, qualitatively different patterns of auditory response are observed for two kinds of acoustic contrasts: (1) the presence or absence of a dip in low-frequency amplitude of a few dB within a voiced interval, and (2) the presence or absence of an abrupt rise in amplitude at higher frequencies, particulary if the onset is accompanied by an initial transient. These contrasting acoustic patterns are similar to those that occur for the sonorant–non-sonorant distinction and for the continuant–non-continuant distinction—two distinctions that appear to be used in the consonant systems of most if not all languages (Maddieson, 1984).

5.3. *Consonantal place of articulation*

One of the basic consonantal distinctions in language is the distinction between labial consonants, produced with a constriction at the lips, and coronal consonants with a constriction formed by the tongue blade. In the case of obstruent consonants that are produced by generating turbulence noise in the vicinity of the constriction, the difference between a coronal and a labial consonant is that there is a short cavity anterior to the constriction for coronals but no such cavity for labials. The acoustic consequences of these different constriction positions are that for coronals the front-cavity resonance gives rise to significant high-frequency spectral energy whereas the spectrum is relatively flat and weak at high frequencies for labials. The theoretical basis for this contrast has been illustrated in Fig. 21. An additional acoustic attribute distinguishing coronals from labials arises from the fact that the tongue body tends to be in a more fronted position for coronals. Consequently the second formant at the implosion or release of the consonant is closer to the third formant for coronals than it is for labials. This is the same property that was discussed earlier in connection with the front–back distinction for vowels.

Manipulation of these parameters such as spectrum of the turbulence noise or the second-formant transition in synthetic consonant-vowel syllables has been shown to elicit identification responses of labial or coronal for various ranges of the parameters (Cooper, Delattre, Liberman, Borst & Gerstman, 1952; Delattre, Liberman & Cooper, 1955). Since more than one acoustic attribute is playing a role in signaling the distinction, it is difficult to isolate the contribution of each attribute without asking listeners to respond to unnatural stimuli in which the different attributes provide conflicting information. Thus, for example, in examining the contribution of the spectrum of the turbulence noise component to the coronal–labial distinction one can adjust the formant transitions to have some neutral pattern intermediate between the patterns for coronals and for labials, but the resulting sound never yields a completely natural coronal or labial consonant.

In one such experiment, synthetic stop-vowel syllables were generated with noise bursts having a range of spectrum amplitudes at high frequencies (Ohde & Stevens, 1983). As expected, listeners identified the initial consonant as /t/ or /d/ for the stimuli with greater high-frequency energy in the burst and as /p/ or /b/ when the high-frequency spectrum amplitude in the burst was weaker. The 50% crossover points in the identification functions occurred when the peak spectrum amplitude of the burst at high frequencies (in the range of F_4 or higher) was approximately 3 dB more than the peak spectrum amplitude in the same frequency range at the onset of the following vowel. Coronals were heard when the relative noise amplitude in this frequency range exceeded 3 dB, whereas labials were heard when the high-frequency noise amplitude in the burst was less than 3 dB above the amplitude at vowel onset. The relative amplitude of the noise burst at the crossover points was shifted a few decibels if the formant transitions were modified to favor the labial or the coronal responses. Discrimination data for stimuli for which the high-frequency spectrum amplitude in the burst is manipulated indicate that there is a peak in discrimination in the region of the continuum where the spectrum amplitude of the burst at high frequencies is 0–5 dB above the spectrum amplitude at the onset of the vowel in the same frequency range (Ohde & Stevens, unpublished data).

An interpretation of these data is that there is a threshold phenomenon in an auditory–acoustic relation such that, except for the factor of 0–5 dB just noted, one type of auditory response is obtained when the high-frequency burst amplitude is weaker than the spectrum amplitude in the adjacent vowel and another type of response is obtained when the burst is stronger. The boundary or threshold between the two types of response occurs about where the high-frequency amplitude at onset does not show an overshoot relative to the following vowel. As has been noted, experimental evidence that can test the validity of this simple picture is clouded by the fact that at least one another acoustic property, corresponding to the front–back distinction, contributes to the coronal–labial distinction.

Another acoustic dimension that is related to consonantal place of articulation is the degree to which the spectrum sampled in the consonantal interval (or at the edge of this interval) is characterized by prominences that are in the midfrequency range (usually in the range of F_2 or F_3) and are comparable in amplitude to and contiguous with spectral prominences in the adjacent vowel. This acoustic attribute is sometimes called *compactness* (Jakobson, *et al.*, 1963). The more a spectral peak is prominent in that it projects above adjacent spectral valleys and is, therefore, well separated from other peaks, the greater is its compactness. The degree of compactness can be decreased either by widening a spectral prominence or by decreasing its amplitude in relation to the corresponding prominence in the adjacent vowel.

Experiments in which these measures are manipulated in a burst that precedes a synthetic vowel show that the initial consonant in the synthetic syllable is heard as a velar stop consonant for the bursts with a single compact spectral prominence and as a labial or alveolar consonant when the compactness is decreased (Hawkins & Stevens, 1987). The shift in identification from velar to non-velar occurs when the peak spectrum amplitude of the prominence is about 3–5 dB below the peak amplitude of the corresponding prominence in the adjacent vowel. Since the burst occurs after an interval of silence, adaptation will tend to enhance the auditory response at the burst onset, so that a burst that is slightly weaker than the adjacent vowel prominence will yield a response in the auditory nerve that is comparable to or exceeds that in the following vowel. What appears to be necessary is that there should be a midfrequency prominence in the spectrum during or immediately adjacent to the consonantal closure interval, and that there should not be an abrupt change in the frequency or amplitude of the prominence as the consonant is released. That is, the prominence continues across the release gesture.

In a similar perceptual experiment with fricative consonants, a continuum of consonant–vowel stimuli ranging from /sa/ to /ša/ was constructed by generating noise with strong high-frequency energy and by systematically raising the amplitude of a prominence in the noise spectrum located at the frequency of the third formant (Stevens, 1985). When the amplitude of this prominence was weak, the consonant was identified as /s/, whereas when it was strong the consonant was heard as /š/. The crossover occurred when the spectrum amplitude of this prominence was equal (within 1 dB) to the amplitude of the F_3 prominence in the adjacent vowel.

This prominence in the spectrum, whether it be in the burst or at voicing onset for a stop consonant, or in the frication noise for a fricative consonant, is in the frequency range of roughly 1–3 kHz, with a bandwidth of 200–300 Hz, i.e., somewhat less than the critical bandwidths indicated in Fig. 25. The estimated bandwidth for this prominence, based on theoretical analysis and a limited set of measurements, is shown as the dotted line in Fig. 25. One might expect, then, that this spectral prominence would give rise to a response pattern in the auditory nerve that shows synchrony to the frequency of the prominence for a range of fibers having characteristic frequencies surrounding this frequency. This pattern would contrast with the pattern obtained from consonants such as alveolars and labials, for which there is no such midfrequency prominence in the spectrum. Chomsky & Halle (1968) have proposed that the feature [anterior] be used to distinguish between velars or palatals (which are [-anterior]) on the one hand, and alveolars and labials on the other.

6. Discussion

We have reviewed a number of articulatory dimensions that give rise to plateau-like regions in relations between acoustic and articulatory parameters, as well as several acoustic dimensions that are characterized by similar relations between auditory and acoustic parameters. These plateau-like regions are separated by regions in which there is a rapid change in the acoustic or auditory parameter, of the type schematized in Fig. 1. We have suggested that these relations could form one basis for establishing the acoustic, auditory, and articulatory correlates of the features that specify phonological distinctions in language. Features whose correlates are based on acoustic–articulatory or auditory–acoustic relations of this kind have the desirable attributes that (1) only limited articulatory precision is required to obtain a desired auditory response when a feature is implemented, and (2) there is a large difference in the acoustic or auditory patterns for articulatory configurations or states corresponding to a particular feature opposition.

Most of the acoustic parameters involved in the acoustic–articulatory or auditory–acoustic relations that we have considered are specified in relational terms. That is, the parameters are defined in terms of a relation between components in different parts of the spectrum, or a relation between spectral components at nearby points in time. Thus, for example, some parameters are specified in terms of the degree of proximity of two spectral peaks, or the enhanced prominence of one spectral peak independent of the frequency of the peak, at least within certain limits. Other parameters are described in terms of the amount of change in the amplitudes of spectral components in particular broad frequency ranges as a vocal-tract constriction is formed or is released. Thus for the most part acoustic parameters are not specified in absolute terms such as the frequency of a spectral peak or the absolute time interval between two acoustic events. If they are measured on an appropriate scale, relational parameters corresponding to particular features are more likely to be independent of vocal-tract size, speaking rate, and phonetic context than are absolute values of parameters such as the frequencies of spectral components.

Our discussion of the various parameters that can lead to quantal relations of the type schematized in Fig. 1 is by no means complete. We have attempted, however, to consider parameters that give insight into the features that play a distinctive role across a variety of languages. These include the features *high*, *low*, *back*, *round*, *nasal*, *breathy voicing*, and *pressed voicing*, which are usually represented in the sound when the vocal tract is relatively open, and *sonorant*, *continuant*, *coronal*, *anterior*, *strident*, and *voiced*, which are represented in the sound when the vocal tract is more constricted. The evidence for quantal relations is not equally strong for all of these features. For example, our understanding of auditory processing in the F_1 region of the spectrum is not sufficiently well advanced that we can make clear statements about the possible role or quantal or threshold phenomena underlying the features *high*, *nasal*, *breathy voicing*, and *pressed voicing*. (Some comments on these questions have been given in Sections 2.5 and 5.1.3.) Another feature that is only partially understood is the feature *continuant*. Further work is needed to specify in detail the nature of the acoustic signal at the release of stop, affricate, and nasal consonants, particularly the release transient for stops and affricates. Perceptual experiments are then needed to determine whether there are threshold effects related to these onset transients that help to distinguish stop from fricative consonants. The details of the acoustic–articulatory relations and the relevant perceptual attributes for consonants (such as retroflex consonants) produced with a front-cavity resonance in the F_3–F_4 region have yet to be worked out (Stevens & Blumstein, 1975; Ladefoged & Bhaskararao, 1983). Similar problems exist for some of the other features in the above lists.

It is natural to ask whether all distinctive features have their bases in quantal relations of the type schematized in Fig. 1, either in the acoustic–articulatory domain or in the auditory–acoustic domain (or both). Or, is it the case that for some features their categorial nature is completely unrelated to peripheral articulatory and auditory mechanisms, and is based solely on more central processes? We have taken as a working hypothesis that quantal relations at the articulatory or auditory level underlie all features. There are, however, several features (other than those just discussed) that at first glance would appear not to fit this pattern.

One such dimension that seems not to be consistent with a quantal interpretation is duration. There is wide variation in the physical length of segments described as *short* and *long* in languages that use duration contrastively (see, for example, Lehiste, 1984). The process by which listeners utilize duration information in the face of variations in speaking rate and context is not understood, and hence it is not yet possible to determine whether auditory processes such as adaptation might play a role in identifying this feature (a possible approach to this question has been proposed by Goldhor, 1985). Other features that present a challenge to the working hypothesis are features that specify tone, the feature *advanced tongue root*, and the feature *distributed* that distinguishes apical from laminal consonants. Furthermore, there are areas of distinctive feature theory where some adjustment or reformulation of the features may be necesssary, and the possible role of quantal relations in these cases will have to be examined.

One consequence of relations like that in Fig. 1 is that there can be differences in the detailed way in which particular features or feature combinations are implemented from one language to another. That is, the value of a particular articulatory or acoustic parameter used to implement a feature in one language may be somewhat different from that in another, but both implementations may lead to an auditory response that is essentially the same, as evaluated along a particular dimension. For example, the high front unrounded vowel /i/ is known to be higher and more fronted in some languages than in others (e.g., Disner, 1983). Or a retroflex stop consonant may be produced in one language with a more posterior point of palatal contact of the tongue tip than in another language (Ladefoged & Bhaskararao, 1983). In the case of the vowel /i/, it is probable that in different languages F_3 and F_4 have the required degree of proximity as specified by the 3.5 Bark criterion (or by some other criterion yet to be established), but the languages may differ in the formant spacing that is used within this range, or they may differ along some other dimension that does not play a distinctive role in forming the contrast in the languages. For the retroflex consonant, both languages presumably adjust the vocal-tract cavities posterior and anterior to the constriction so that there is proximity of F_3 and F_4 at the consonant boundary. The actual frequencies of F_3 and F_4 or their degree of proximity may be different, however. For both of these examples, it is expected that the different implementation of the segments would be audible to speakers of the different languages, but the difference would not be sufficiently great to shift the sounds out of the designated regions of the relevant auditory–articulatory relations. Detailed study of these cross-language differences would constitute a useful test of the adequacy of some of the proposed quantal relations.

As a consequence of relations of the type shown in Fig. 1, we are proposing that certain ranges of acoustic and articulatory parameters are preferred over others in establishing the inventory of phonetic features that are used in language. Each of the features has a different acoustic correlate, and we can imagine a simultaneously implemented[2] bundle of features, or a segment, as being represented as a point or a region in multidimensional acoustic space, where the dimensions represent the acoustic parameters corresponding to the different features. The premise of this paper, then, is that the inventory of sounds in any language will have a preference for particular regions of this acoustic space.

Needless to say, any given language uses only a small subset of the possible combinations of features. A detailed discussion of the principles that underlie the selection of this subset is outside of the scope of this paper. In the case of vowels, Liljencrants & Lindblom (1972) have proposed a principle of maximum perceptual separation, and have presented convincing evidence in support of this principle based on an examination of the vowel systems in a number of languages. They assume that, at least in the two-dimensional formant space considered in their paper, the vowels can take on a continuous range of positions. The quantal relations discussed here would suggest, however, that there is a tendency toward particular regions within this space. The dimensions F_1 and F_2 in the space used by Liljencrants & Lindblom may not be the dimensions that are most appropriate for representing vowels, since, as we have seen, relational properties (such as formant spacing or degree of prominence of a spectral peak) may come closer to capturing the perceptually relevant properties. Fant (1973), among others, has proposed some alternative dimensions for representing vowels.

In any event, the Liljencrants–Lindblom analysis indicates that, for vowel systems, a principle of maximum perceptual saliency or maximum perceptual contrast is consistent with their data. Such a principle will tend to impose rounding on non-low back vowels, since rounding will result in values of F_2 that are lower and closer to F_1, and hence will provide greater separation of back vowels from front vowels. In the case of front vowels, unrounding will lead to higher frequencies for F_2 and/or F_3, and again will increase the perceptual contrast with back vowels. Vowel systems that are relatively sparse, then, will tend to contain rounded non-low back vowels and unrounded front vowels. This conclusion can only be reached, however, if the front–back dimension is regarded as being more salient perceptually than, say, the rounding dimension.

Similar principles presumably apply to consonant systems. Thus, for example, one might consider the feature *continuant* to be one of the more salient features, because the auditory system is known to produce a distinctive response pattern to sounds with an abrupt onset (Delgutte & Kiang, 1984*b*). The abruptness of the onset for an obstruent unaspirated stop consonant is presumably stronger if the consonant is voiceless than if it is voiced, because of the greater intraoral pressure for voiceless stops. Thus voiceless stop consonants are to be preferred over voiced stops because of the stronger representation of the feature *continuant*. Similar observations can be made with respect to other feature combinations, indicating that a feature can be represented in the sound more strongly when it is implemented with certain other features.

The point of these comments with regard to the present paper is that the relations of the type schematized in Fig. 1 do not lead to equally salient contrasts between regions I and III for all features. For some articulatory and acoustic parameters the contrasts are inherently more salient than for others. Furthermore, the strength of the perceptual

[2] The nature of the articulatory and acoustic correlates of some features (particularly those for consonants) requires that the acoustic manifestations of the features occur within a narrow time interval. For other features, however, the acoustic implementation is not necessarily tied to particular landmarks in the signal.

change as the articulatory parameter takes on a range of values across region II may be dependent on other parameters that co-occur with the parameter under study. (Further discussion of these questions appears in Stevens & Keyser, 1989.)

We conclude by summarizing some of the questions that have been raised in this discussion as we attempt to examine in greater depth the notion that auditory and articulatory constraints play a significant role in shaping the phonetic categories that are used in language.

(1) As a broader range of phonetic dimensions is examined, will it turn out that all distinctive features have their bases in relations like those in Fig. 1, either in the acoustic–articulatory domain or in the auditory–acoustic domain? An answer to this question may have to await further experimental data in auditory physiology, as researchers begin to investigate patterns of response of units in the cochlear nucleus and higher levels in the auditory system, as well as a deeper understanding of the physiological and acoustical aspects of speech production.

(2) Examination of the acoustic and articulatory representations of speech sounds across languages indicates some differences in these representations for segments that appear to be classified by the same features. These observations naturally lead to the question: what acoustic differences have the potential of signaling a phonetic contrast, and what acoustic differences are not sufficient to produce a phonetic contrast but can occur consistently across languages and can be heard by speakers of those languages?

(3) Phonetic contrasts that may be based on auditory–articulatory relations of the type discussed here probably differ in their perceptual salience. Further work is needed to determine which articulatory and acoustic parameters lead to the more salient contrasts, and the conditions under which the strength of a particular contrast can be enhanced.

This paper is dedicated to the memory of Dennis Klatt. The ideas in this paper were influenced by discussions with a number of individuals, including Sheila Blumstein, Gunnar Fant, Morris Halle, Jay Keyser, Dennis Klatt, Sharon Manuel, and Joe Perkell. The influence of these colleagues, as well as a number of other members of the Speech Group at MIT, is gratefully acknowledged. Preparation of this paper was supported in part by grants from the National Institute of Neurological and Communicative Disorders and Stroke (NS-04332) and from the National Science Foundation (8418733-BNS).

References

Alwan, A. (1986) Acoustic and perceptual correlates of pharyngeal and uvular consonants. Unpublished SM thesis, Massachussetts Institute of Technology, Cambridge MA.

Bickley, C. (1982) Acoustic analysis and perception of breathy vowels. Speech Communication Group Working Papers I, MIT, Cambridge MA, pp. 71–81.

Bothorel, A., Simon, P., Wioland, F. & Zerling, J.-P. (1986) Cinéradiographie des Voyelles et Consonnes du Français. *Travaux de l'Institute de Phonétique de Strasbourg*, Strasbourg.

Carlson, R., Fant, G. & Granström, B. (1975) Two-formant models, pitch, and vowel perception. In *Auditory analysis and perception of speech* (G. Fant & M. A. A. Tatham, editors), pp. 55–82, London: Academic Press.

Carlson, R., Granström, B. & Fant, G. (1970) Some studies concerning perception of isolated vowels. *Speech Transmission Laboratory QPSR2-3*, Royal Institute of Technology, Stockholm, pp. 19–35.

Chistovich, L. A. & Lublinskaya, V. V. (1979) The "center of gravity" effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli, *Hearing Research*, **1**, 185–195.

Chistovich, L. A., Sheikin, R. L. & Lublinskaya, V. V. (1979) "Centres of gravity" and spectral speaks as the determinants of vowel quality. In *Frontiers of speech communication research* (B. Lindblom & S. Öhman, editors), pp. 143–157. London: Academic Press.

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

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M. & Gerstman, L. J. (1952) Some experiments on the perception of synthetic speech sounds, *Journal of the Acoustical Society of America*, **24**, 597–606.

Delattre, P. C., Liberman, A. M. & Cooper, F. S. (1955) Acoustic loci and transitional cues for consonants, *Journal of the Acoustical Society of America*, **27**, 769–773.

Delgutte, B. & Kiang, N. Y-S. (1984*a*) Speech coding in the auditory nerve: I. Vowel-like sounds, *Journal of the Acoustical Society of America*, **75**, 866–878.

Delgutte, B. & Kiang, N. Y-S. (1984*b*) Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics, *Journal of the Acoustical Society of America*, **75**, 897–907.

Di Benedetto, M.-G. (1987) An acoustical and perceptual study on vowel height. Unpublished PhD thesis, University of Rome.

Disner, S. F. (1983) Vowel quality: The relation between universal and language specific factors. *UCLA Working Papers in Phonetics*, No. 58, University of California at Los Angeles.

Fant, G. (1960) *Acoustic theory of speech production.* The Hague: Mouton.

Fant, G. (1972) Vocal tract wall effects, losses, and resonance bandwidths. *Speech Transmission Laboratory QPSR 2-3*, Royal Institute of Technology, Stockholm, 28–52.

Fant, G. (1973) *Speech sounds and features*. Cambridge MA: MIT Press.

Fujimura, O., & Lindqvist, J. (1971) Sweep-tone measurements of vocal-tract characteristics, *Journal of the Acoustical Society of America*, **49**, 541–558.

Goldhor, R. (1985) Representation of consonants in the peripheral auditory system: A modeling study of the correspondence between response properties and phonetic features. *RLE Technical Report 505*, Massachusetts Institute of Technology, Cambridge MA.

Harnad, S. (editor) (1987) *Categorical perception.* Cambridge: Cambridge University Press.

Hary, J. W. & Massaro, D. W. (1982) Categorical results do not imply categorical perception, *Perception and Psychophysics*, **32**, 409–418.

Hawkins, S. & Stevens, K. N. (1985) Acoustic and perceptual correlates of the nonnasal–nasal distinction for vowels, *Journal of the Acoustical Society of America*, **77**, 1560–1575.

Hawkins, S. & Stevens, K. N. (1987) Perceptual and acoustical analysis of velar stop consonants. In *Proceedings of the eleventh international congress of phonetic sciences*, vol. 5, pp. 342–345, Tallinn.

Jakobson, R., Fant, G. & Halle, M. (1963) *Preliminaries to speech analysis.* Cambridge MA: MIT Press.

Johnson, D. H. (1980) The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones, *Journal of the Acoustical Society of America*, **68**, 1115–1122.

Kakita, Y. & Fujimura, O. (1977) Computational model of the tongue: A revised version, *Journal of the Acoustical Society of America*, **62**, (S1), S15 (A).

Kiang, N. Y-S., Watanabe, T., Thomas, E. C. & Clark, L. F. (1965) *Discharge patterns of single fibers in the cat's auditory nerve*. Cambridge MA: MIT Press.

Klatt, D. H. (1987) Acoustic correlates of breathiness: First harmonic amplitude, turbulence noise, and tracheal coupling, *Journal of the Acoustical Society of America*, **82** (S1), S91 (A).

Klatt, D. H. & Stevens, K. N. (1969) Pharyngeal consonants. *RLE Quarterly Progress Report*, No. 93, Massachusetts Institute of Technology, pp. 207–215.

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

Ladefoged, P. (1983) The linguistic use of different phonation types. In *Vocal fold physiology: contemporary research and clinical issues* (D. M. Bless & J. H. Abbs, editors), pp. 351–360. San Diego: College-Hill Press.

Ladefoged, P. & Bhaskararao, P. (1983) Non-quantal aspects of consonant production: A study of retroflex consonants, *Journal of Phonetics*, **11**, 291–302.

Lehiste, I. (1984) The many linguistic functions of duration. In *New directions in linguistics and semiotics* (J. E. Copeland, editor), pp. 96–122. Rice University Studies, Houston TX.

Liljencrants, J. & Lindblom, B. (1972) Numerical simulation of vowel quality systems: the role of perceptual contrast, *Language*, **48**, 839–862.

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

Ohde, R. N. & Stevens, K. N. (1983) Effect of burst amplitude on the perception of stop consonant place of articulation, *Journal of the Acoustical Society of America*, **74**, 706–714.

Perkell, J. S. (1969) *Physiology of speech production.* Cambridge MA: MIT Press.

Perkell, J. S., Boyce, S. E. & Stevens, K. N. (1979) Articulatory and acoustic correlates of the [s–š] distinction. *Journal of the Acoustical Society of America*, **65**, (S1), S24 (A).

Rosen, S. & Howell, P. (1987) Auditory, articulatory, and learning explanations of categorical perception in speech. In *Categorical perception* (S. Harnad, editor), pp. 113–160. Cambridge: Cambridge University Press.

Shadle, C. (1985) The acoustics of fricative consonants. *RLE Technical Report 506*, Massachusetts Institute of Technology, Cambridge MA.

Stevens, K. N. (1972) The quantal nature of speech: Evidence from articulatory-acoustic data. In *Human communication: a unified view* (E. E. David, Jr & P. B. Denes, editors), pp. 51–66. New York: McGraw-Hill.

Stevens, K. N. (1985) Evidence for the role of acoustic boundaries in the perception of speech sounds. In *Phonetic linguistics: essays in honor of Peter Ladefoged* (V. A. Fromkin, editor), pp. 243–255. London: Academic Press.

Stevens, K. N. (1987) Interaction between acoustic sources and vocal-tract configurations for consonants. In *Proceedings of the eleventh international conference of phonetic sciences*, vol. 3, pp. 385–389. Tallinn.

Stevens, K. N. (1988) Modes of vocal-fold vibration based on a two-section model. In *Vocal physiology: voice production mechanisms and functions* (O. Fujimura, editor), pp. 357–371. New York: Raven Press.

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

Stevens, K. N. & House, A. S. (1955) Development of a quantitative description of vowel articulation, *Journal of the Acoustical Society of America*, **27**, 484–493.

Stevens, K. N. & House, A. S. (1956) Studies of formant transitions using a vocal tract analog, *Journal of the Acoustical Society of America*, **28**, 578–585.

Stevens, K. N. & Keyser, J. (1989) Primary features and their enhancement in consonants, *Language*, **65**, 81–106.

Stevens, K. N. Andrade, A. & Viana, C. (1987) Perception of vowel nasalization in VC contexts: A cross-language study, *Journal of Acoustical Society of America*, **82** (S1), S119(A).

Stevens, K. N., Fant, G. & Hawkins, S. (1987) Some acoustical and perceptual correlates of nasal vowels. In *In honor of Ilse Lehiste: Ilse Lehiste Pühendusteos* (R. Channon & L. Shockey, editors), pp. 241–254. Dordrecht, Holland: Foris Publications.

Stevens, K. N., Keyser, J. & Kawasaki, H. (1986) Toward a phonetic and phonological theory of redundant features. In *Invariance and variability in speech processes* (J. S. Perkell & D. H. Klatt, editors), pp. 426–449. Hillsdale NJ: Erlbaum.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M. & Öhman, S. E. G. (1969) Cross-language study of vowel perception, *Language and Speech*, **12**, 1–23.

Syrdal, A. K. (1985) Aspects of a model of the auditory representation of American English vowels, *Speech Communication*, **4**, 121–135.

Syrdal, A. K. & Gopal, H. S. (1986) A perceptual model of vowel recognition based on the auditory representation of American English vowels, *Journal of the Acoustical Society of America*, **79**, 1086–1100.

Traunmüller, H. (1981) Perceptual dimension of openness in vowels, *Journal of the Acoustical Society of America*, **69**, 1465–1475.

Young, E. D. & Sachs, M. B. (1979) Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers, *Journal of the Acoustical Society of America*, **66**, 1381–1403.

Zwicker, E. (1961) Subdivision of the audible frequency range into critical bands (Frequenzgruppen), *Journal of the Acoustical Society of America*, **33**, 248.