Submitted to Handbook of Phonetic Sciences, W. Hardcastle and J. Laver (eds.)

Articulatory-Acoustic-Auditory Relationships

Kenneth N. Stevens Research Laboratory of Electronics and Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139

1 Introduction

The process of communication by means of speech involves the generation of sound by a speaker and interpretation of that sound by a listener. In preparation for production of sound, the speaker plans the utterance in a linguistic form, one component of which is a concatenation of words that are organized into phrases and larger units. The words are represented in the lexicon as structured sequences of segments each of which is characterized by discrete classes or features. The linguistic representation then initiates commands to the muscles that are responsible for respiration and for manipulating the various laryngeal and supraglottal structures. The sound that reaches the ear of the listener is processed by the peripheral auditory system and the output of this processing forms the basis for lexical access and ultimately interpretation of the utterance.

One of the central problems of research in speech science is to model the relation between the continuously varying speech signal produced by a speaker and the discrete linguistic representation in terms of words, segments, and features. Of particular interest are two kinds of discreteness: discreteness in time (e.g., the acoustic representation of the word *sick* has three segmental units), and discreteness in terms of phonological categories (e.g., *sick* contrasts with *seek* in the vowel segment and with *sit* in the final consonant segment).

The exercise of examining the relations between articulation and sound and between speechlike sounds and auditory responses can potentially shed light on these questions.

The sound produced by the vocal tract can be described in terms of a number of parameters such as relative frequencies of formants, descriptors of the waveform of glottal excitation, amplitude and spectrum of turbulence noise, fundamental frequency changes, etc. These parameters change as the positions and states of the various articulators are manipulated. As has been discussed elsewhere (Stevens, 1972; 1989), the changes in the acoustic parameters as the articulatory parameters vary through ranges of values are often not monotonic. The relation between an acoustic parameter and an articulatory parameter often shows a region in which the acoustic parameter is relatively stable, and an adjacent region where there is an abruptness or discontinuity in the relation. The same type of relation is also often observed between certain acoustic parameters and some aspects of the auditory response to the different acoustic patterns. That is, as an acoustic parameter is manipulated, there are abrupt changes in the auditory response for certain values of the parameter.

These acoustic-articulatory and auditory-acoustic relations provide some basis for the two types of discreteness noted above. As the positions of the articulators change with time, there will be points in time where acoustic discontinuities or dislocations occur, and other points where there are extrema (maxima or minima) in some acoustic parameters such as formant frequencies. These points in time can be regarded as markers for segmental units. The quantal acoustic properties in the vicinity of these discontinuities or extrema help to define the features of these segmental units.

This view of the role of acoustic-articulatory-auditory relations in helping to define the correlates of segmental units provides motivation for the discussion in this chapter.

2 Articulatory-acoustic relations

The production of speech sounds can be described as a process involving (1) the generation of sound sources, and (2) the filtering of these sound sources by the airway above the glottis (Fant, 1960). The sound sources are usually created by forming a narrowing of the airway at or above the glottis, and causing a rapid flow of air through the constriction. The generation of sound can occur through three different mechanisms: modulation of airflow through vocal-fold vibration, turbulent airflow near a constriction, and generation of a transient due to rapid release of pressure in the vocal tract. Filtering of the sound sources can be described in terms of transfer functions relating the spectrum of the vocal-tract output to the spectrum of the sources. This transfer function is usually dominated by peaks at the natural frequencies of the vocal tract, particularly the natural frequencies of the portion of the vocal tract that is downstream from the sources, or frequencies determined by upstream portions of the tract that are closely coupled to the sources.

In the following sections we describe the characteristics of the various sources and of the vocal-tract filtering that shapes the sources to produce the sound pressure at a distance from the speaker.

2.1 Sound sources in the vocal tract

2.1.1 Sources at the glottis

When the vocal folds are positioned sufficiently close together and a subglottal pressure is applied, vibration of the vocal folds is initiated. The changing shape of the folds during a cycle of vibration is illustrated by the succession of coronal sections in Fig. 1 (Baer, 1975). The cycle begins with the folds approximated at their superior edges. The subglottal pressure produces abducting forces on the lower surfaces of the folds (shown by the arrows in panel 1 of the figure), and causes these surfaces to displace laterally. This lateral movement of the inferior surfaces eventually causes the upper edges to move apart, re-

sulting in airflow through the glottis. When this airflow occurs, the pressure in the glottis decreases, and consequently there is a reduction in the lateral forces on the lower surfaces. The restoring force due to the vocal-fold stiffness then causes these surfaces to reverse direction and to displace inwards (panels 4 and 5 of Fig. 1). The lower surfaces eventually come together and the airflow is cut off, and these surfaces are once again subjected to the abducting force from the subglottal pressure. The upper edges come together somewhat later in the cycle. The cycle of vibration then begins again. Maintenance of vibration requires a certain minimum subglottal pressure that is about 3 cm H_2O (Titze, 1992).

The airflow through the glottis during a cycle of vibration has a waveform like that shown in Fig. 1. The volume velocity waveform is skewed somewhat to the right because the mass of the air in the glottis and in the subglottal and supraglottal airways causes the volume velocity to lag behind the area change (Rothenberg, 1981). This lag causes the airflow to increase more slowly during the opening phase and to decrease more rapidly during the closing phase, as the figure shows.

The frequency of glottal vibration is determined primarily by the stiffness and mass of the folds, although the subglottal pressure can also have a small influence on the frequency. The principal mechanism for changing the frequency is to stretch the vocal folds by increasing their length. This stretching is accomplished through contraction of the cricothyroid muscle. The frequency can also be influenced by contraction of the vocalis muscle which is embedded within the vocal folds, but the relation between vocalis muscle contraction and frequency is less direct (Hirano et al., 1969).

The periodic pulses of volume velocity, schematized in Fig. 2a, form a source of excitation of the vocal tract during glottal vibration. The spectrum of these pulses has the form shown in Fig. 2b. At high frequencies, above about 500 Hz, the amplitudes of the harmonics decrease as $1/f^2$ (f = frequency) if there is an abrupt discontinuity of the slope

of the waveform at the instant of closure. At frequencies below 500 Hz, the amplitudes of the harmonics decrease less rapidly with increasing frequency.

When the glottis is adjusted to a relatively open position by spreading the arytenoid cartilages apart, the vocal folds will no longer vibrate. In this configuration, the aerodynamic forces cannot provide sufficient energy to maintain vibration. The flow through the open glottis will, however, be turbulent, and noise is generated as a consequence of this turbulent flow. Noise produced in this manner is called aspiration noise.

This noise source can have two components: (1) a component that can be modelled as a sound-pressure source distributed over a region up to 3 cm downstream from the glottis, where the glottal airstream impinges on the ventricular folds and on the surface of the epiglottis; and (2) a component that can be modelled as a volume-velocity source due to fluctuations in the flow through the glottis. The component that is modeled as a sound-pressure source is in series with the impedance Z_b looking back from the source toward the glottis. This impedance Z_b is that of a short transmission line terminated in the glottal impedance. This sound-pressure source in series with Z_b can be represented as an equivalent volume-velocity source in parallel with Z_b, and can be combined with the second volume-velocity component to form an equivalent noise source. The estimated spectrum of this equivalent noise source is shown in Fig. 3 (Stevens, in preparation). For different glottal areas A_g and subglottal pressures P_s, this noise spectrum is scaled up or down in proportion to $P_s^{1.5} A_g^{0.5}$, for a reasonable range of P_s and A_g observed in normal speech production (Stevens, 1971).

During modal vocal-fold vibration, with a periodic waveform and spectrum shown in Figs. 1 and 2, it is expected that some turbulence noise will be generated in the flow. If we assume an average glottal area of 0.04 cm^2 during modal vibration, then the estimated spectrum of this equivalent volume-velocity noise source is as shown by the lower

solid line in Fig. 4. The spectrum of the periodic source for modal vibration is shown for comparison (upper solid line), and is well above the spectrum of the noise source for frequencies up to 5 kHz.

When the vocal folds are partially abducted at the vocal processes, vibration can still be maintained, but the waveform is modified, as schematized in Fig. 5a. The modified waveform differs from the waveform for modal vibration in several ways. The average flow is larger, and, since the vocal folds do not come together during the cycle, the airflow does not have a sharp change in slope when it reaches a minimum flow. The net effect of this abducted configuration on the spectrum of the flow is shown in Fig. 5b. The periodic component of the spectrum is reduced at high frequencies, and the amplitude of the noise component is increased, as shown by the dashed lines in Fig. 4. The net result is that the noise spectrum can become comparable to the periodic component at middle and high frequencies. This mode of glottal vibration is called breathy voicing. When the glottis is spread still more, vocal-fold vibration will cease, as noted above, and only the noise component will remain, as in Fig. 3.

Adduction of the vocal folds by contraction of the lateralis muscle can produce a sufficiently large compression force that the subglottal pressure cannot push the folds apart, and no sound is generated at the glottis. For a somewhat smaller adducting force, a condition of pressed voicing can occur. The glottal volume-velocity waveform for this configuration has narrower pulses than for modal vibration, and a more abrupt termination of each pulse. The spectrum has a somewhat greater amplitude at high frequencies and a lower amplitude for the first one or two harmonics than the spectrum in Fig. 2 for modal vibration. The component due to turbulence noise is negligible.

When a narrow constriction is formed in the airway above the glottis, during the production of a consonant, pressure can build up behind the constriction, assuming that

the glottis is not completely adducted. For an obstruent consonant produced in this manner, there is a reduced pressure drop across the glottis. Vocal-fold vibration can be maintained only if this pressure drop does not decrease below a certain critical value. Continued vibration can be facilitated for an interval of a few tens of milliseconds if the volume of the vocal tract behind the constriction is actively expanded by lowering the larynx and/or advancing the tongue root, thereby preventing the supraglottal pressure from decreasing too rapidly (Rothenberg, 1968; Westbury, 1979).

2.1.2 Frication noise

When a sufficiently narrow constriction is formed in a region of the vocal tract above the glottis, there is a possibility of turbulence noise generation in the vicinity of the constriction when there is airflow through the constriction. The amplitude, spectrum, and location of this sound-pressure source depend upon the pressure drop across the constriction, the cross-sectional area of the constriction, and the locations of any obstacles in the airstream downstream from the constriction. The volume velocity U through the constriction is related approximately to the cross-sectional area A and the pressure drop ΔP by the equation

$$\Delta P = \frac{\rho U^2}{2A^2}, \qquad (1)$$

where ρ = density (van den Berg et al., 1957). This equation is valid within about 20 percent for most constriction shapes, as long as the area A is not too small (i.e., not less than about 0.01 cm^2).

The amplitude of the turbulence noise source for a given airflow through the constriction is greatest if the air stream impinges directly on an obstacle. This condition is approximated during the production of a fricative consonant like /s/ or /š/. The amplitude and spectrum of the sound-pressure source for this situation has been estimated from measurements with mechanical models (Shadle, 1985; Pastel, 1987; Stevens, 1993). This

spectrum is shown in Fig. 6, for a mechanical model in which air flows through a tube containing a narrow constriction. For the upper spectrum, the volume velocity through the constriction of 420 cm^3/s and the cross-sectional area of the constriction of 0.08 cm^2. For a different volume velocity U and cross-sectional area A, this spectrum should be scaled up or down in proportion to $U^3A^{-2.5}$, as illustrated for a lower volume velocity in Fig. 6. Also, when there is not an obstacle directly in the airstream, the amplitude of the noise can be reduced by up to 10-15 dB. As the figure shows, the spectrum of the noise source has a broad peak in the mid-frequency range, and decreases at high frequencies. As will be shown later, this source is filtered by the airway, particularly the part of the airway downstream from the constriction.

2.1.3 Transient sources

When a complete closure is made by an articulator in the oral cavity and there is an increased or decreased pressure behind the constriction, rapid release of the constriction causes an acoustic transient to be generated at the instant of release. This transient is a consequence of the abrupt airflow as the compressed air behind the constriction is released. It occurs at the release of stop consonants (Maeda, 1987; Massey, 1994), and is particularly salient at the release of certain clicks (Ladefoged and Traill, 1994) and affricate consonants. The transient can be modeled as an abruptly changing volume-velocity source at the constriction, with a brief duration of less than 1 ms.

2.2 Vowels and sonorant consonants

There is a large class of speech sounds that are produced with a sound source at or near the glottis and with no buildup of pressure due to airflow through a narrow constriction in the airway above the glottis. For these sounds the properties of the sound source at the glottis are relatively uninfluenced by the configuration of the vocal tract above the glottis. Differentiation between sounds produced in this manner is achieved either by manipulating the properties of the glottal source in the manner described in Section 2.1

above, or by manipulating the filtering of the source through adjustment of the shape of the airways above the glottis. Two types of filtering can be distinguished: (1) with a vocal tract shape that is relatively unconstricted in the oral or pharyngeal region, and (2) with a relatively narrow constriction or closure in the oral region. The first class consists of the vowels, and the second class encompasses the sonorant consonants, including liquids, glides, and nasal consonants. For the consonant [h], the vocal tract above the glottis may also be relatively unconstricted, but the glottis is spread. There is a much larger airflow than for vowels, with consequent turbulence noise generation near the glottis. For all of these sounds, the filtering can be described by a transfer function, which is the ratio of the amplitude of the acoustic volume velocity at the mouth (or at the nose, or both) to the volume velocity of the source, as a function of frequency.

2.2.1 Nonnasal vowels

When the velopharyngeal port is closed, the transfer function for vowels can usually be approximated by an all-pole function (Fant, 1960). The poles of the transfer function are the natural frequencies of the vocal tract, or the formants. A typical transfer function, for a vowel with a uniform vocal-tract cross-sectional area, as it might be produced by an adult female, is shown in Fig. 7. The formants are manifested as peaks in the transfer function.

The frequencies of the formants, particularly the first two formants $F1$ and $F2$, are dependent on the shape of the airway between the glottis and the lips, and this shape in turn is determined by the position of the tongue body and the lips. For purposes of determining its acoustic behavior, the shape of the airway is specified by the cross-sectional area as a function of the distance from the glottis, called the area function. The average spacing of the lowest three or four formants is dependent on the overall length ℓ of the vocal tract, and is given approximately by $\frac{c}{2\ell}$, where c is the velocity of sound. The third and higher formant frequencies tend to be less sensitive to tongue-body position than the

first two formant frequencies. When the tongue body and lips are positioned such that the cross-sectional area of the vocal tract is uniform, the frequencies of the formants are approximately equal to $\frac{c}{4\ell}$, $\frac{3c}{4\ell}$, $\frac{5c}{4\ell}$, ...

The relation between tongue-body and lip position on the one hand and the formant frequencies on the other is a complex one, but there are some general principles governing this relation. One approach to examining this articulatory-acoustic relation is through a perturbation analysis (Chiba and Kajiyama, 1941; Schroeder, 1967; Fant, 1980). For a given area function, each natural frequency or formant can be characterized by a distribution of the amplitude of sound pressure and volume velocity over the length of the vocal tract. When the area function is modified by making a small perturbation of cross-sectional area over a local region of the vocal tract, a given natural frequency is displaced upward or downward by an amount that depends on the location of the perturbation in relation to the maxima in these distributions. In particular, when a small decrease in cross-sectional area is made in a region where the distribution of volume velocity is a maximum for a formant (or the distribution of sound pressure is a minimum), then the frequency of that formant will be displaced downward. Likewise if the decrease is in a region where there is a minimum in the volume-velocity distribution, then the shift in the formant is upward. When the cross-sectional area is increased rather than decreased in these local regions, the formant shifts in the opposite direction.

Application of these principles to a vocal tract with a uniform cross-sectional area is illustrated in Fig. 8. The area function is shown at the top left, and below it are the distributions of volume-velocity and sound-pressure amplitude for the first two formants. The volume velocity is constrained to be a maximum at the open end and a minimum at the closed (glottis) end. To the right are curves showing the relative amounts of shift of the first three formants ($\Delta F1$, $\Delta F2$ and $\Delta F3$) when a small local decrease in cross-sectional area is made at different points along the length of the uniform tube. A local increase

in the cross-sectional area will, of course, lead to values of $\Delta F1$, $\Delta F2$ and $\Delta F3$ that are opposite in sign to the changes resulting from a local decrease in area. One can observe a symmetry in these functions: a given perturbation in the posterior half of the vocal tract yields formant shifts that are equal and opposite to those obtained with a similar perturbation at a symmetrically located point in the anterior half of the tract. Alternatively, it is noted that perturbations of opposite sign at symmetrically located points enhance the formant shifts produced by one such perturbation.

These perturbation functions can be used to infer how the formants shift as a result of different movements of the tongue body and lips (Mrayati et al., 1988). For example, narrowing of the area function over the posterior half of the length of the vocal tract and widening over the symmetrical anterior half gives rise to an increase in the first-formant frequency. In terms of tongue-body position, this change in shape (relative to a uniform vocal tract) is achieved by displacing the tongue body downward. This movement is assisted by lowering the jaw. On the other hand, narrowing of the area function in the anterior half of the vocal tract and widening it in the posterior half causes the first-formant frequency to decrease. This change in the area function is accomplished by raising the tongue body in the oral cavity. For both of these movements — tongue-body lowering and tongue-body raising — there are limits to the amount of movement consistent with maintaining an airway that is not too constricted.

Similar interpretations can be made for perturbations of the vocal-tract shape that arise from front-back movements of the tongue body and from rounding of the lips. For example, if the vocal tract is narrowed in a region of the hard palate, Fig. 8 shows that the second-formant frequency increases. A fronting movement of the tongue body gives rise to this type of change in the area function. Likewise, narrowing in the pharyngeal region causes a decrease in the second-formant frequency. When the narrowing is at the lips, the second-formant frequency also decreases, and this decrease is enhanced by a con-

comitant narrowing in the posterior part of the oral cavity. These types of perturbation are achieved by displacing the tongue body in a posterior direction.

Estimates of the acoustic consequences of changes in vocal-tract shapes through perturbation analysis leads, then, to following general relation between tongue-body position and the first two formant frequencies: high or low tongue-body positions lead, respectively, to low or high $F1$; front or back tongue-body positions lead, respectively, to high or low $F2$. In addition, the low $F2$ with a high back tongue-body position is enhanced by lip rounding. These relations are summarized in schematic form in Fig. 9, which is a plot of $F2$ versus $F1$. The four combinations of high and low $F1$ and $F2$ are marked by points at the corners of a quadilateral, and the area functions that give rise to these combinations are superimposed on the plot. These area functions are shown as concatenations of uniform tubes with different lengths and cross-sectional areas.

These four vowels represent extremes in the vowel space in that it is not possible to form vowel configurations for which the formant frequencies lie outside of the quadrilateral defined by the four points in Fig. 9. For the front vowels, the tongue-body position is adjusted so that second-formant frequency has a maximum value that is as close to $F3$ as possible, consistent with maintaining a vowel-like configuration. On the other hand, for the back vowels, the tongue-body position and lip rounding are adjusted to produce a minimum value of $F2$. When the tongue body is positioned so that $F2$ has a maximum or minimum, it is expected that $F2$ would be relatively insensitive to small perturbations in tongue-body position (Stevens, 1972; 1989).

Other vowels that may be distinctive in a given language have formant frequencies that lie either on the sides of the quadrilateral in Fig. 9 or inside the quadrilateral. The $F1$ and $F2$ values for some of the vowels in English are plotted in Fig. 10 (Peterson and Barney, 1952). Also shown in a separate panel of the figure is a plot of $F2$ versus $F3$

for the same vowels. This part of the figure shows that the range of $F3$ values across the vowels is considerably narrower than the $F2$ range. Rounded vowels ([u ʊ ɔ]) have a somewhat lower $F3$ than unrounded vowels, and $F3$ is higher for the palatalized vowel [i] than for the other front vowels. The frequencies in Fig. 10 are average values for adult female speakers. They would be 10-20 percent lower for adult male speakers on the average, because the overall vocal tract length is usually greater for males than for females.

The modelling of a vowel as a quasiperiodic volume-velocity source filtered by an all-pole transfer function is an approximation that accounts for the principal attributes of vowel spectra under most conditions. There are, however, several situations in which this simple model requires some modification. These include: (1) the existence of an opening through the velopharyngeal port to the nasal cavity; (2) a shaping of the tongue blade and tongue body so that there is more than one acoustic path around or under the tongue; and (3) a glottal opening that is large enough to provide acoustic coupling to tracheal resonances. The first and second of these influences will be considered in later sections of this chapter. The effect of acoustic coupling to the trachea is to introduce pole-zero pairs into the transfer function. These pole-zero pairs appear as additional peaks and valleys in the spectrum of a vowel. The additional peaks occur in the vicinity of resonances of the subglottal system. For adult speakers, the first three of these resonances are in the frequency ranges 600-800 Hz, 1400-1800 Hz, and 2000-2500 Hz (Ishizaka et al., 1976; Cranen and Boves, 1987; Klatt and Klatt, 1990).

Figure 11 gives examples of spectra of vowels (produced by female speakers) in which these spectral perturbations due to tracheal coupling are apparent. The arrows indicate the approximate locations of the subglottal resonances. In some spectra a sharp minimum is apparent adjacent to an extra peak, but this evidence for a zero is not always clear.

The tracheal resonance that is evident most consistently in the examples in Fig. 11

and in other examples is the one in the range 1400-1800 Hz. When the second vocal-tract resonance $F2$ becomes close to this tracheal resonance or passes through it, the tracheal resonance can cause a perturbation in the amplitude and frequency of the $F2$ prominence, particularly for speakers who have particularly strong evidence for a tracheal resonance. This tracheal resonance provides a natural boundary in the $F2$ value between back and front vowels.

2.2.2 Sounds produced with an aspiration noise source

When there is an aspiration noise source in the vicinity of the glottis, it has been noted that this source can be represented as an equivalent volume-velocity source, with a spectrum as shown in Fig. 3. The transfer function for this source is essentially the same as that for the periodic glottal source, except that there are modifications in the formant bandwidths and frequencies due to the partially open glottis. The bandwidths, particularly for the lowest two or three formants, may be considerably wider than those for modal glottal vibration.

The spectrogram of the utterance [əhet], shown in Fig. 12, illustrates the differences in the spectrum for a given vocal-tract configuration for the two types of sources. The spectra below the spectrogram show the differences in the source spectrum as well as differences in the bandwidths of the lower formants. The first formant is highly damped when there is aspiration because of the large acoustic losses at and below the glottis in this low-frequency range. At high frequencies, in the range of $F4$ and $F5$, the spectrum amplitude of the aspiration noise (left panel) is comparable to the spectrum amplitude in the adjacent interval where the source is from glottal vibration (middle panel). Immediately after onset of glottal vibration, there is breathy voicing, with a reduced amplitude of the second harmonic relative to the first harmonic, and a widened bandwidth of $F1$ compared with that in the middle of the vowel (right panel).

2.2.3 Transitions into and out of consonants

Production of a consonant is usually achieved by forming a narrow constriction in the oral portion of the vocal tract. This constriction is made with one of three different articulators: the lips, the tongue blade, or the tongue body. The narrowing can result in complete closure of the airway or just a partial closure. The changes in vocal-tract shape that occur as this constriction is formed or is released give rise to changes in the formant frequencies, called formant transitions. The pattern of movements of the different formants depend upon which articulator forms the constriction, how the articulator is shaped, and where the articulator is placed. Spectrograms illustrating the different patterns of change of formant frequencies for different voiced stop consonants are given in Fig. 13.

The formation of a consonantal constriction in the oral cavity with the lips, the tongue blade or the tongue body always causes the first-formant frequency to decrease. This observation can be derived from the perturbation concepts described above, since a narrowing is in the anterior portion of the vocal tract where there is a maximum in the velocity distribution for $F1$. The movements of the second and third formants depend where in the oral cavity the constriction is formed, as illustrated in Fig. 13. The shifts in $F2$ and $F3$ for different places of articulation are shown schematically in Fig. 14 when the unconstricted vocalic configuration has a uniform area function. A labial constriction is formed by narrowing the vocal tract at the anterior end, and this narrowing is accompanied by a raising of the mandible, causing a modest tapering of the area function over the anterior part of the vocal tract, as the figure shows. The $F2$ and $F3$ values for this configuration are lower than the values for the idealized vowel. An alveolar constriction, which is formed by raising the tongue blade and raising the mandible, results in an increase in $F2$ and $F3$, since the constriction is formed in a region where there is a maximum in the sound-pressure distribution for both $F2$ and $F3$. Both $F2$ and $F3$ are resonances of the cavity behind the constriction for an alveolar consonant. The schematized shape for this consonant is given in the upper right portion of Fig. 14. A velar constriction is made

by raising the tongue body, and for the uniform tube the narrowing occurs in a region where $F2$ is displaced upward (near a maximum in sound pressure) and $F3$ is displaced downward (near a minimum in sound pressure), as shown in the figure. The arrows in Fig. 14 depict the direction of movement of $F2$ and $F3$ when the consonant configuration is released into the neutral vowel (although the shape of this trajectory is not necessarily the linear shape schematized in the figure).

The act of creating the consonantal constriction with a particular articulator allows some freedom in the shaping of regions of the vocal tract that are not directly involved in producing the constriction (Ohman, 1966). These portions of the vocal tract can be manipulated in anticipation of segments adjacent to the consonant, such as the following vowel. For example, if a labial stop consonant is produced in a syllable with a following front vowel, the tongue body can be displaced forward during the time the lips are closed, so that the vocal-tract shape is different from the schematized shape for a labial consonant in Fig. 14. The values of $F2$ and $F3$ for the labial configuration are shifted relative to the values in Fig. 14 for a following neutral vowel. Likewise, if the labial consonant precedes a back vowel, the tongue body can be shifted back during the consonant closure, in anticipation of the following vowel.

Shown in Fig. 15 are the estimated starting $F2$ and $F3$ frequencies for labial, alveolar, and velar stop consonants as they occur immediately preceding eight different vowels in American English. In each part of the figure, the vowel labels identify $F2$ and $F3$ values for the steady-state vowels, as shown above in Fig. 10. In each panel the estimated $F2$ and $F3$ values at the release of one of the consonant types are joined to the vowel into which the consonant is released, with the arrow indicating the direction of movement from consonant to vowel.

The values of $F2$ and $F3$ for consonants in Fig. 15 are estimated from several sources

of data, and have been adjusted to show consistent monotonic behavior across front and back vowels (Stevens et al., 1966; Kewley-Port, 1982; Sussman et al., 1991). The frequencies are intended to represent the natural frequencies of the vocal tract when it is in the consonantal position preceding the different vowels, and they may be slightly different from the measured average values of $F2$ and $F3$ during the first one or two glottal pulses following the consonantal release. For each consonant, the $F2$ and $F3$ values are enclosed by separate contours for the three groups of vowel contexts: front vowels ([i ɪ ɛ æ]), back unrounded vowels ([ɑ ʌ]), and back rounded vowels ([u ʊ]).

For labial consonants, the range of starting frequencies for $F2$ before different vowels is about 800 Hz. When a labial consonant precedes a back vowel, the starting frequencies of $F2$ and $F3$ are similar to the formant frequencies for the vowels; that is, the transitions of the formants are small. When a labial consonant is produced before a front vowel, the $F2$ starting frequency is significantly below that for the vowel, and $F3$ is also lower for the most fronted vowels [i] and [ɪ]. In general, the starting frequency for $F2$ is in the range normally occupied by back vowels, so that when there is a following front vowel, $F2$ must cross over the region separating back from front vowels. In the case of the vowel [i], the movement of the point in the $F2$-$F3$ plane following the release follows a curved trajectory, reflecting the fact that the front cavity resonance at release is $F2$, but this resonance shifts to $F3$ as the lip opening increases. The back-cavity resonance remains fixed during this maneuver; it is $F3$ at the time of release and then shifts to being affiliated with $F2$.

The range of starting frequencies is more constrained for alveolar consonants (Fig. 15b). Apart from the high front vowel [i] and the high back rounded vowel [u], the consonantal $F2$ and $F3$ values are relatively tightly clustered. When an alveolar precedes a back vowel, $F2$ falls, and passes through the region that separates front from back vowels. The starting frequencies for velar consonants (Fig. 15c) cluster into three groups depending on whether the vowel is a front vowel, a back unrounded vowel, or a back rounded vowel. For

velar consonants before back vowels, $F2$ falls and $F3$ rises, whereas when the following vowel is a front vowel both $F2$ and $F3$ tend to fall.

When a stop consonant closure is formed following a vowel, the transitions of the formants are usually similar to (but not always identical to) those for a consonant-vowel sequence. However, these transitions at the consonant closure can also be influenced by the vowel that follows the consonant.

The formant transitions for fricative consonants produced with the same labial, alveolar, and velar constrictions are similar to those shown in Fig. 15, although the transitions tend to be less extreme, i.e., the formant starting frequencies are closer to $F2$ and $F3$ for the vowel than they are for stop consonants. When the place of articulation for a consonant produced with the tongue blade is different from alveolar, the $F2$ and $F3$ starting frequencies will be shifted somewhat from the values given in Fig. 15b.

The patterns of formant movements for different consonant-vowel combinations in Fig. 15 are based on data from English, but are expected to be characteristic of languages for which there is no contrast in tongue-body position for consonants, such as palatalization, velarization or pharyngealization. When such a contrast does exist, it is expected that the $F2$ and $F3$ starting frequencies for a given consonant will be much more constrained (Ohman, 1966).

When the vocal-tract constriction for a consonant is formed in the pharyngeal region, the perturbation principles discussed in Section 2.2.1 predict a pattern of formant movements quite different from that for consonants with a constriction in the oral region of the vocal tract. The most salient difference is that the first-formant frequency does not decrease when a pharyngeal constriction is formed, since such a constriction is made in the posterior half of the vocal-tract area function (Klatt and Stevens, 1969). A constriction

in the lower pharyngeal region is expected to cause a greater increase in $F1$ than a constriction in the uvular region. Typical values of $F2$ and $F3$ for these consonants adjacent to the three vowels [i ɑ u] are given in Fig. 16 (Alwan, 1986). Pharyngeal consonants ([ħ] in the figure) show a rising $F3$ into the following vowel whereas uvulars ([χ]) do not, and both pharyngeals and uvulars have a rising $F2$ when the vowel is [i].

2.2.4 Nasal vowels and consonants

When the soft palate is lowered to create a velopharyngeal opening, the acoustic coupling to the nasal cavity causes modifications in the vocal-tract transfer function and hence in the spectrum of a vowel. Since there are substantial individual differences in the morphology of the nasal cavity, the effects of this nasal coupling or the vowel spectrum can be quite variable, particularly in the middle and high-frequency range. At lower frequencies, however, there are several acoustic consequences of nasalization that are more consistent across different speakers, although all of these attributes may not be evident in all cases. One such effect is an increased bandwidth of the first formant that occurs because of greater acoustic losses in the nasal cavity, which has a large surface area covered with mucosa. Another acoustic consequence of nasalization is the introduction of additional peaks in the spectrum, due to pole-zero pairs in the transfer function (Dang et al., 1994). One such peak usually occurs in the frequency range 800-1100 Hz (Stevens, forthcoming), depending on the area of the velopharyngeal opening. Another peak has been observed at lower frequencies in nasal vowels in French, usually below the first formant (Delattre, 1954), and this could be caused by a resonance of the maxillary sinuses (Maeda, 1982a) or by enhanced low-frequency energy in the glottal spectrum.

These three acoustic manifestations of nasalization for a vowel are schematized in the spectra in Fig. 17. The solid line is a hypothetical spectrum envelope in the lower frequency range for a nonnasal vowel. The dashed line shows how the spectrum envelope would be modified by each of these effects: widened $F1$ bandwidth (leading to a reduced

prominence of the $F1$ peak), additional pole-zero pair due to the nasal cavity proper, and an enhancement at low frequencies. The overall effect of these individual spectral modifications is to flatten the spectrum in the vicinity of the first formant (Maeda, 1982b). This spectrum flattening is the result of widening the $F1$ bandwidth and "filling in" the spectrum above and below $F1$ so that the prominence of $F1$ as an isolated peak is lessened.

The spectrograms and spectra in Fig. 18 illustrate the contrasting acoustic properties of nasal and nonnasal vowels in French. In each of the two words *combat* and *engage*, the first vowel is a nasal vowel and the second is nonnasal. The spectra of the two nasal vowels (top spectra) show an enhanced first harmonic and a greatly increased first-formant bandwidth, compared with the nonnasal vowels (bottom spectra). Evidence for an additional resonance in the vicinity of 1000 Hz is obscured, since $F2$ for the vowel is also in this frequency range.

A nasal consonant is produced by making a complete closure with one of the articulators, while maintaining an open velopharyngeal port. Sound is radiated from the nose. During the closure interval, the transfer function from glottal volume velocity to nose volume velocity contains poles that are the natural frequencies of the combined vocal and nasal tracts, together with zeros at frequencies for which the impedance looking into the oral tract from the region of the velopharyngeal opening is zero. Typically, the lowest resonance is around 250-300 Hz (for an adult speaker) and the next one is in the vicinity of 800-1000 Hz. This latter resonance is called the nasal resonance, since it is due primarily to the nasal cavity. The next highest pole is at a frequency that is approximately equal to the second formant that would occur for a nonnasal consonant configuration, since the effect of nasal coupling on the second and higher formants is relatively small. This frequency depends upon the place of articulation of the consonant and on the following vowel, as shown in Fig. 15. The spectral prominence due to this formant may be obscured, however, by the presence of a zero that is nearby in frequency. The frequency

of this zero is roughly equal to $\frac{c}{4\ell_f}$, where ℓ_f is the length of the front cavity from the velopharyngeal opening to the point of oral closure. This frequency is expected to be in the range 1000-1200 Hz for a labial consonant and 1500-1800 Hz for an alveolar consonant.

These attributes in the frequency range up to about 2000 Hz are evident in the spectra shown in Fig. 19. Three pairs of spectra are given, corresponding to the three nasal consonants [m] (in *a mill*), [n] (in *a knock*), and [ŋ] (in *sing out*). Also shown is a spectrogram of *a knock*. The spectrum in the upper panel in each case is sampled in the nasal murmur just before the release into the vowel, and the lower spectrum is sampled in the vowel just after the consonant release. The low-frequency first resonance is evident in all the nasal murmurs. The nasal resonance is labeled as FN. The true second formants $F2$ and $F3$ are also labeled, and these frequencies are roughly in accord with the value on the chart in Fig. 15. The movement of $F2$ and $F3$ following the release of [n] can be seen in the spectrogram. Evidence for the zero in the spectrum of the nasal murmur can be seen at about 1100 Hz for [m], 1500 Hz for [n] and around 2100 Hz for [ŋ], although the last of these is not well-defined.

A salient acoustic attribute at the release of a nasal consonant is the abrupt increase in spectrum amplitude in the middle and high-frequency range as the output shifts abruptly from the nose to the mouth, and the frequency of the zero shifts rapidly downward. As Fig. 19 shows, the jump in spectrum amplitude in the frequency range of $F2$ is about 20 dB, when upper and lower spectra are compared. This abrupt amplitude change is due to the rapid change in the first-formant frequency at the release as well as to the abrupt downward shift in the zero.

2.2.5 Liquids and glides

Like the vowels, the liquids and glides are produced with a source at the glottis, but with a constriction in the airway that is relatively narrow. The constriction is not so narrow

however, that it creates a significant pressure drop when the vocal folds are vibrating with a modal configuration. The constriction is narrow enough, however, that the airflow creates a resistance that increases the acoustic losses and hence increases the bandwidth of certain formants. As a consequence, when the vocal tract is in the most constricted configuration for such a consonant the spectrum has some peaks that are less salient then others. The formant with the increased bandwidth is usually $F2$ or $F3$. The constriction also has the potential effect of loading the glottal source, resulting in a reduced amplitude of the source and a greater downward tilt of the source spectrum at high frequencies (Bickley and Stevens, 1986). This influence on the source, combined with the low first-formant frequency that occurs when the vocal tract is constricted in the oral region, causes liquids and glides to have a low-frequency spectrum amplitude that is reduced relative to that of an adjacent vowel. Thus there are three general properties of liquids and glides that distinguish them from vowels: a reduced low-frequency spectrum amplitude, an additional decrease in amplitude at high frequencies, and a reduced prominence of the second or third formant peak. All of these attributes are not necessarily present in all instances of these sounds. In addition to these properties, the spectra for liquids ([l] and [r]) have some irregularities at high frequencies (in the $F3$ range) that distinguish them from the glides ([w] and [j]). These irregularities arise from additional pole-zero pairs in the transfer function because of multiple acoustic paths around the constriction formed by the tongue blade.

Examples of spectra sampled at the most constricted point during each of the liquids and glides in American English are compared with the spectra sometime later in the following vowel, in the top and bottom panels in Fig. 20. In each case, the consonant is produced in a symmetrical intervocalic position with the vowel selected to have $F1$ and $F2$ values similar to those for the consonant. In this way, a comparison can be made between the vowel and consonant spectra with minimal influence of the formant frequencies on the spectrum shapes. Comparison of the upper and lower panels indicates how the

spectrum changes from the more constricted region for the consonant to the more open vowel configuration. For example, there is rapid increase in spectrum amplitude in the $F2$ region and at higher frequencies for both [l] and [r], with a high-frequency $F3$ for [l] and a lower $F3$ for [r].

The glides [w] and [j] are produced with vocal-tract configurations similar to those for the high vowels [u] and [i], except that the tongue-body height (and, in the case of [w], the degree of lip rounding) is more extreme. For the back glide [w], the narrowed airway causes an increase in the bandwidth of $F2$. The high-frequency amplitude for [j] is considerably lower than that for the following vowel, presumably due to an increased tilt in the glottal spectrum.

2.3 Obstruent consonants

In Section 2.2, we have described some articulatory-acoustic relations when the configuration of the vocal tract is such that the airflow does not cause a significant pressure drop in the supraglottal airway. For those configurations, the acoustic source is always at or near the glottis, and the pressure drop across the glottis is equal to the subglottal pressure. When the constriction in the airway above the glottis is sufficiently narrow, a pressure drop can occur across the constriction. The increased pressure behind the constriction causes a decreased pressure drop across the glottis, and the transglottal pressure becomes less than the subglottal pressure. This is the condition that exists when an obstruent consonant is produced.

An obstruent consonant has two acoustic attributes that distinguish it from sounds produced with no pressure drop in the vocal tract above the glottis: (1) because of the reduced transglottal pressure, the strength of vocal-fold vibration is reduced or vocal-fold vibration ceases, and (2) turbulence noise is generated due to the rapid airflow at the constriction. If the size of the constriction is maintained at a small nonzero value over

an interval of time of a few tens of milliseconds, the noise source continues throughout this interval, and a fricative consonant is generated. If a complete closure is formed to produce a stop consonant, then turbulence noise is produced at the constriction for a brief interval of time after the release. A transient source may also occur at the time of release of the consonant.

In the next two sections we describe the articulatory, aerodynamic and acoustic events for the most common voiceless stop and fricative consonants. We then consider how this description is modified when different laryngeal adjustments are made during the production of these consonants. The transitions of the formants when these consonants are produced with different articulators have been discussed above in Section 2.2.5, and we concentrate here on the time variation of the sources and the spectrum of the sound resulting from the frication and transient sources.

2.3.1 Stop consonants

The production of a stop consonant consists of two events: a closure and a release of an articulator. If the consonant is in intervocalic position, there are formant transitions as the closure is being formed and after the release has occurred. We assume in this section that the vocal folds remain in a state appropriate for modal voicing for a vowel, and that there are no active adjustments of this state when the consonant is produced. The sequence of mechanical and aerodynamic events when an intervocalic stop consonant is produced are summarized in Fig. 21. The time variation of the cross-sectional area of the constriction formed by the articulator is shown in part (a) of the figure: the cross-sectional area decreases rapidly, the closure remains for 80-odd ms, and then the area increases. Immediately after closure occurs, the intraoral pressure increases (also shown in Fig. 21a) and the pressure across the glottis decreases. As the intraoral pressure increases, the force from this pressure causes the walls of the vocal tract and of the glottis to displace outwards. The outward movement of the walls U_w, together with the flow U_g through the

glottis and U_c through the supraglottal constriction, is plotted in Fig. 21b. These curves depict average flows, and do not show fluctuations due to glottal vibration. The actual glottal flow is schematized in part (c) of the figure. When the transglottal pressure drops below a threshold value of about 3 cm H_2O, vocal-fold vibration ceases. The decrease in transglottal pressure is sufficiently rapid that only one or two glottal pulses occur following the consonant closure.

After a few tens of milliseconds, the intraoral pressure becomes equal to the subglottal pressure. At the time when the closure is released, there is a rapid outward flow of air through the constriction. This airflow occurs in two successive phases: an initial brief transient flow as the compressed air is released, and a longer peak in flow as the vocal-tract walls displace back to their rest position and as there is resumption of airflow through the glottis. Vocal-fold vibration begins when the transglottal pressure increases above a threshold value of about 3 cm H_2O. These aerodynamic events following the release result in a sequence of acoustic sources: (1) a transient source, (2) a brief turbulence noise source (frication noise) in the vicinity of the constriction, (3) a possible brief interval of aspiration noise at the glottis, and (4) onset of the quasiperiodic volume-velocity source at the glottis. This sequence of acoustic sources is shown schematically in Fig. 22.

The detailed aerodynamic and acoustic events depend on the rate of closure or release of the articulator forming the stop consonants. The rates of change of cross-sectional area shown in Fig. 21 are typical of labial and alveolar stop consonants. The rates for velar consonants are slower, and the slower release leads to a longer burst and a longer time interval from the release to the onset of glottal vibration.

These various sources are filtered by the vocal tract, and the output shows spectral peaks, with the degree of prominence of each peak depending on the spectrum and location of the source. The transient and frication noise sources excite primarily the resonances of

the part of the airway that is downstream from the constriction, whereas the aspiration or periodic source at the glottis excites all of the vocal-tract resonances.

The waveform and spectrogram for the utterances ['bɑpə], shown in Fig. 23, illustrate some of the acoustic attributes of an unaspirated labial stop consonant [p] as it occurs before a reduced vowel. In this example, there is a time interval of about 12 ms from the consonant release (shown by an arrow on the waveform) to the onset of the first glottal pulse. The falling second-formant transition as the consonantal closure is formed, and the rising $F2$ transition after the labial release are evident in the spectrogram.

The spectrum sampled over a brief time interval following the release of the burst of the unaspirated consonant [p] in ['bɑpə] is shown in Fig. 24a. Overlaid on this spectrum is the spectrum of the vowel averaged over a 4-ms interval in the second period of the following vowel. The spectrum of the burst has no major spectral prominences, reflecting the fact that there is no acoustic cavity anterior to the constriction to filter the noise and transient sources at the constriction. Minor spectral peaks are evident in the consonant spectrum as a consequence of weak excitation of the vocal tract behind the consonant constriction. Because there is no enhancement of the source by a front-cavity resonance, the burst spectrum for [p] is relatively weak, and is lower in amplitude than the spectrum peaks in the following vowel.

Similar pairs of spectra of the consonant burst and the following vowel (second glottal period) are shown for unaspirated [t] and [k] in Figs. 24b and 24c. In the case of the alveolar consonant [t], there is a major spectral peak in the frequency range 3500-4000 Hz, due to a resonance of the short cavity anterior to the constriction formed by the tongue tip. The amplitude of this spectrum peak is 10-15 dB greater than the amplitude of the spectrum peak that is in the same frequency range for the vowel. In the case of [k] (Fig. 24c), the major spectrum peak is at 1500 Hz, reflecting a resonance of a front

cavity with a length of about 6 cm. The amplitude of the peak is about equal to the amplitude of the $F2$ spectrum peak in the following vowel. A second resonance of this cavity is evident at about 4500 Hz. The location of the tongue-body constriction for [k] is dependent on the following vowel, and hence the vowel influences the frequency location of the major mid-frequency peak in the spectrum of the burst. When [k] is followed by a front vowel, the frequency of the spectral prominence for the burst is usually in the vicinity of $F3$ for the vowel, rather than $F2$, as shown in the figure.

Theoretical studies of models, together with data of the type given in Fig. 24, show that the spectrum of the burst at the release of a stop consonant has distinctive characteristics depending on which articulator produces the consonantal closure and where that articulator is positioned. The spectrum of the burst has no major prominences for a labial stop, has a peak in the $F4$ or $F5$ range for an alveolar stop, and has a prominence in the $F2$ or $F3$ region for a velar stop.

2.3.2 Fricative consonants

When a fricative consonant is produced, the cross-sectional area of the constriction becomes narrow enough that continuous turbulence noise is generated in the vicinity of the constriction. The time variation of the cross-sectional area when a fricative consonant is in intervocalic position has the form shown in Fig. 25a. When the fricative consonant is voiceless, it is necessary to spread the glottis in order to provide sufficient airflow to generate the frication noise. A typical time course of the glottal opening for the intervocalic fricative is also given in Fig. 25a. The time variation of the supraglottal and glottal constriction areas are shown by two sets of lines. The solid lines are the areas that would exist if there were no increase in intraoral pressure. The increased intraoral pressure creates forces on the surface of the supraglottal articulator and on the surface of the glottis, causing an increase in the two areas. The calculated modified area traces are shown by the dashed lines. The pressures and flows resulting from this pattern of change

of the areas can be calculated approximately using equation (1), with the result shown in Figs. 25b and 25c.

As Fig. 25b shows, the transglottal pressure decreases as the constriction is formed and increases at the release, and consequently there is cessation of glottal vibration during this time interval. Frication noise is generated throughout the time when there is an increased intraoral pressure.

An example of the fricative [s] in an intervocalic context is shown in the spectrogram in Fig. 26. The transitions of the first and second formants into and out of the consonant are evident, as is the frication noise and the cessation of glottal vibration. The spectrum of the frication noise is displayed below the spectrogram, together with the spectrum of the vowel shortly after the onset of glottal vibration. The frication noise has a spectrum similar to that of the [t] burst in Fig. 24b, with a broad high-frequency peak. This peak is in the $F5$ range here, and is about 20 dB higher in amplitude than the corresponding peak in the adjacent vowel.

Typical spectra for several fricatives in English are shown in Fig. 27. These spectra have contrasting shapes as the place of articulation changes from labiodental to dental to alveolar to palato-alveolar. The labiodental and dental fricatives have no major spectral prominence, at least up to 8 kHz, since there is basically no front-cavity resonance. There are only small differences between these two spectra, and the spectrum amplitude at high frequencies (in the $F4$ and $F5$ range) is below the spectrum amplitude of the adjacent vowel in the same frequency range. For the palatoalveolar fricative [š], there are spectral peaks corresponding to $F3$ and $F4$, at 2600 and 3300 Hz, and there is some noise excitation of $F2$. The $F3$ and $F4$ peaks are resonances of the narrow passage formed by the tongue blade and of the cavity in front of the constriction formed by the anterior surface of the tongue blade.

2.3.3 Voicing for obstruents

The articulatory-acoustic relations discussed above for stop and fricative consonants assume glottal configurations that are appropriate for a voiceless unaspirated stop consonant or for a voiceless fricative consonant. When these glottal configurations are combined with the constricting gestures of one of the articulators in the oral cavity, vocal-fold vibration is inhibited during most of the time interval in which the constriction is formed. Various adjustments of the glottal state and/or of the pharyngeal volume can be used to facilitate glottal vibration during the constricted interval or to introduce aspiration noise at the glottis either before the constriction is formed or after it is released.

For example, vocal-fold vibration will continue throughout the closure interval for a stop consonant if the transglottal pressure is maintained at a sufficiently high value, assuming that the glottis is not strongly abducted or adducted. This condition is achieved by actively expanding the vocal-tract volume in the pharyngeal region, either by advancing the tongue root or by lowering the larynx, or both. For a fricative consonant, glottal vibration is maintained if the glottis is only partially spread, and can also be facilitated if the pharyngeal volume is actively expanded. An example of a voiced stop consonant in intervocalic position is given in Fig. 28a. The spectrogram shows the continuing glottal vibration throughout the closure interval. The spectra sampled in the closure interval and in the adjacent vowel show the drop in amplitude of the first-formant prominence and the significant reduction in amplitude at high frequencies due to the attenuation of sound through the vocal-tract walls and neck in the high-frequency range. The spectrum of the burst, in the bottom panel of Fig. 28a, again shows the relatively weak amplitude and flat spectrum characteristic of a labial (as in Fig. 24a).

A voiceless aspirated stop consonant is normally produced in prevocalic position by spreading the glottis during the closure interval, so that the glottal opening is a maximum at the time of release of the consonant. In the 60-odd ms following the release, the glottis

returns to a modal configuration, and aspiration noise is generated at the glottis during the initial few tens of ms. In some languages, voiceless aspirated stop consonants contrast with voiced aspirated stops, in which glottal vibration continues through much of the closure interval, and then the glottis is spread to cause aspiration noise to be generated following the release.

Spectrograms and spectra of voiceless and voiced aspirated stop consonants are shown in Figs. 28b and 28c, together with spectra sampled at selected points in the utterances. The spectrum of the aspiration noise in [p] (middle spectrum) is similar to that for [h], discussed earlier in Section 2.2.2, in that it has prominent formant peaks (in contrast to the spectrum at onset). For the voiced aspirated consonant, there is greater low-frequency energy in the aspiration. In both consonants, the bottom spectra show that there is a time interval when breathy voicing occurs, in which the amplitudes of the second and higher harmonics are reduced in relation to the amplitude of the first harmonic. This attribute of breathy voicing was noted in Section 2.1.1 and in Fig. 2.5.

3 Acoustic-auditory relations

3.1 Vowels

In Section 2.2.1 we observed that the pattern of formant frequencies for vowels depends on the position of the tongue body and the configuration of the lips. To produce a pattern with a high first-formant frequency, the tongue body must be low, and to produce a low $F1$ the tongue body is in a raised position in the oral cavity. Likewise, the second-formant frequency is high and close to $F3$ when the tongue body is in a fronted position, and is low and distant from $F3$ when the tongue body is backed. For nonlow vowels, this attribute of a low $F2$ is enhanced if the lips are rounded.

This classification of vowels as high or low or as front or back also appears to be

grounded in basic auditory responses to stimuli that contain spectral prominences similar to those for vowels. Chistovich and her colleagues (Chistovich and Lublinskaya, 1979; Chistovich et al., 1979) carried out a series of experiments in which listeners adjusted the frequency of a single-formant stimulus so that there was a match in quality to a two-formant stimulus (with frequencies F_a and F_b). Various parameters of the two-formant stimulus were manipulated, including the frequency spacing F_b-F_a between the formants and the relative amplitudes of the two spectral prominences. When the spacing between the two formants was greater than about 3.5 Bark, listeners adjusted the frequency of the one-formant matching stimulus to be equal to F_a or F_b, or else they showed great variability in setting the frequency of the matching stimulus, depending on the relative amplitudes. A different pattern of response was obtained when F_b-F_a was less than about 3.5 Bark. The matching formant was adjusted to F_b when the amplitude of that spectral prominence was substantially greater than that of F_a, and vice versa. However, when the amplitudes of the two prominences were similar (within a range of about 30 dB), the matching frequency was intermediate between F_a and F_b. In this case, then, the listeners tended to match to the "center of gravity" of the two-formant stimulus rather than to one formant or the other. Based on these experiments, one can conclude that when two spectral prominences are separated by less than about 3.5 Bark, there is some kind of auditory integration of the two prominences, whereas each prominence maintains a separate auditory representation when the separation is greater than 3.5 Bark.

These results of auditory matching experiments help to provide an auditory basis for the classification of vowels. For example, Syrdal and Gopal (1986) used the data from Peterson and Barney (1952) to show that the $F3$-$F2$ values for vowels from a wide range of American English speakers are less than 3 Bark for front vowels and more than 3 Bark for back vowels. Their analysis of the Peterson and Barney data also showed that vowels separated into high and nonhigh categories according to whether $F1$-$F0$ was less than or greater than 3 Bark. Evidence for the relevance of $F1$-$F0$ (in Bark) as a normalized

measure of vowel height has been reported by Traunmüller (1981).

Further evidence that vowel quality for front vowels is determined by the "center of gravity" of the complex of higher formants ($F2$-$F3$-$F4$) comes from experiments reported by Carlson et al. (1970). They showed that a vowel that is judged to have a quality similar to a multi-formant front vowel can be synthesized with a two-formant vowel in which $F1$ for both vowels are the same. To obtain a match in vowel quality, subjects placed the second formant ($F2'$) of the two-formant synthetic vowel at a frequency that was usually between $F2$ and $F3$, and was sometimes between $F3$ and $F4$, depending on the values of $F2$ and $F3$ in relation to $F4$. However, when the spacing between $F2$ and $F3$ was greater than 3.0 Bark (as it is for back vowels), $F2'$ was set equal to $F2$ for the multi-formant vowel.

3.2 Nasalization

As has been shown in Section 2.2.4, one of the ways that a vowel spectrum can be modified to produce a class of sounds that contrasts with nonnasal vowels is through nasalization. The various acoustic consequences of acoustic coupling to the nasal cavity through the velopharyngeal port have been summarized in Section 2.2.4. Several experiments have examined the perceptual consequences of introducing these acoustic attributes into synthetic utterances. For example, it has been shown that listeners judge a vowel to be nasal when (1) the spectrum amplitude at low frequencies is enhanced (Delattre, 1954); (2) the bandwidth of the first formant is increased (Hawkins and Stevens, 1983; Chen, 1994); and (3) a pole-zero pair is introduced to create an additional prominence in the vowel spectrum above the first formant (Hawkins and Stevens, 1985; Chen, 1994).

Maeda (1982b) has pointed out that all of these acoustic consequences of forming a velopharyngeal opening for a vowel contribute to a single global acoustic property: flattening the spectrum in the frequency range extending up to about 1500 Hz. (See

also Stevens, 1985a.) The nature of this spectral flattening for nasal vowels has been illustrated in Fig. 17. Maeda proposed a metric for quantifying the degree of "flatness" of the spectrum in this frequency range, but noted that some refinement of this metric is necessary.

3.3 Distinction between aspirated and unaspirated stop consonants

As has been discussed in Section 2.3.3, the distinction between voiced and voiceless stop consonants in prestressed position is marked in part by a difference in the time from the consonant release to onset of glottal vibration (voice-onset time, or VOT), this time being greater for the voiceless cognate. When the VOT is manipulated in a synthetic consonant-vowel syllable to be greater than about 25 ms, the consonant is heard as voiceless, and when it is less than 25 ms it is heard as voiced. Experiments with nonspeech stimuli consisting of two sounds with different onset times have examined the ability of listeners to judge the temporal order of the two sounds (Hirsh, 1959; Pisoni, 1977). These experiments have shown that a difference in onset times of at least 20 ms is needed for listeners to reliably judge the temporal order and to perceive the two onsets as successive rather than simultaneous events. One can conclude from these results with nonspeech stimuli that this 20-ms limitation is a basic property of the perceptual system, and this property appears to be exploited in establishing a system of phonetic contrasts between voiceless stop consonants with a longer VOT and unaspirated stop consonants with a short VOT.

3.4 Place for consonants

When a consonant is produced with a narrow constriction in some region of the oral cavity, the transitions of the formants when the consonant is released into the following vowel are different for different places of articulation for the consonant, as has been discussed in Section 2.2.3. If the consonant is an obstruent, the spectrum of the frication noise that

is produced also varies with the place of articulation, as illustrated in Figs. 24 and 27. A number of experiments have examined listener responses to synthetic consonant-vowel syllables in which these formant transitions and noise spectra were manipulated. The stimuli that elicit coronal responses (i.e., responses of *t* or *d* to stop consonants and *s* to fricative consonants) tend to be those for which the spectrum amplitude at high frequencies (range of $F4$ or higher) near the vowel onset is greater than or equal to that just after the vowel onset (Blumstein and Stevens, 1980; Ohde and Stevens, 1983; Stevens, 1985b). Labial responses are obtained when the spectrum amplitude at high frequencies in the consonant noise is weaker than that in the following vowel. Responses of *k* or *š* (sometimes classified as *compact* consonants) are obtained when the spectrum of the noise has a narrow prominence in the $F2$ or $F3$ range, with an amplitude that is comparable to that of the corresponding formant peak in the following vowel. These acoustic attributes for labial, alveolar, and velar stop consonants have been illustrated in Fig. 24.

A general conclusion from these and other results is that the identification of place of articulation for an obstruent consonant in a CV syllable is determined by the spectrum of the noise portion and by the nature of the spectrum change at the transition into the vowel (Stevens, 1985b). A *compact* consonant has a spectrum with a midfrequency prominence that is narrower than an auditory critical band, whereas *diffuse* consonants like *p* or *t* do not. The labial and coronal consonants can usually be distinguished on the basis of whether or not there is a broad high-frequency prominence with an amplitude that is comparable to or greater than the high-frequency spectrum amplitude in the following vowel.

4 Summary

In this chapter we have given a number of examples that show how certain acoustic characteristics of the radiated sound change as the articulatory parameters describing the

vocal-tract shapes are manipulated through a range of values. In some of these examples, an abrupt change or discontinuity in the acoustic attributes occurs when a consonant is produced by forming or releasing a narrow constriction in the vocal tract. This abrupt change may be due to a reduction in the amplitude of the glottal source, or to the rapid introduction of a frication noise source, or to a switching of the output from the mouth to the nose. In other examples involving obstruent consonants, adjustments in the positioning and shaping of the constriction can cause quantal changes in the vocal-tract resonance that receives major excitation from the frication noise source. Examples are the spectra of the stop bursts and fricative noise in Figs. 24 and 27. In the case of vowels, there appear to be tongue-body and lip positions for which the second formant achieves a relatively stable maximum or minimum value that is only weakly sensitive to the positioning of the tongue body.

Examination of the perception of sounds with speechlike properties shows that as certain dimensions of the stimuli are manipulated through a range of values, listener responses change in a discontinuous manner. These dimensions include the spacing between two formants or between $F1$ and $F0$, spectral "flattening" in the first-formant region, the timing of a stop consonant release and onset of glottal vibration, and the prominence of spectral peaks in frication noise in relation to those of an adjacent vowel.

It appears that the non-monotonicity of these articulatory-acoustic and auditory-acoustic relations forms one basis for the design of a set of phonetic categories that are used in languages. Discontinuities or abruptnesses in these relations are utilized to form consonantal landmarks in the signal, and acoustic parameters in the vicinity of the landmarks provide cues for some of the features of the consonant. In the case of vowels, both articulatory-acoustic and auditory-acoustic relations play a role in defining front-back and height distinctions and the nasal-nonnasal distinction.

REFERENCES

1. Alwan, A. (1986) *Acoustic and perceptual correlates of pharyngeal and uvular consonants.* S.M. thesis, Massachusetts Institute of Technology, Cambridge, MA.

2. Baer, T. (1975) *Investigation of phonation using excised larynxes.* Ph.D. thesis, Massachusetts Institute of Technology, Cambridge MA.

3. Bickley, C.A. and K.N. Stevens (1986) *Effects of a vocal-tract constriction on the glottal source: experimental and modelling studies.* **J. Phonetics 14**, 373-382.

4. Blumstein, S.E. and K.N. Stevens (1980) *Perceptual invariance and onset spectra for stop consonants in different vowel environments.* **J. Acoust. Soc. Am. 67**, 648-662.

5. Carlson, R., B. Granstrom and G. Fant (1970) *Some studies concerning perception of isolated vowels.* **Speech Transmission Laboratory Quarterly Progress and Status Report 2-3**, Royal Institute of Technology, Stockholm, 19-35.

6. Chen, M. (submitted) *Acoustic parameters of nasalized vowels in hearing-impaired and normal-hearing speakers.* **J. Acoust. Soc. Am.**

7. Chiba, T. and M. Kajiyama (1941) **The vowel: Its nature and structure.** Tokyo:Tokyo-Kaiseikan.

8. Chistovich, L.A. and V.V. Lublinskaya (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.

9. Chistovich, L.A., R. Sheikin, and V. Lublinskaya (1979) *'Centers of gravity' and spectral peaks as determinants of vowel quality.* In B. Lindblom and S. Ohman (eds.) **Frontiers of Speech Communication Research**, London: Academic Press, 143-157.

10. Cranen, B. and L. Boves (1987) *On subglottal formant analysis.* **J. Acoust. Soc. Am. 81**, 734-746.

11. Dang, J., K. Honda, and H. Suzuki (1994) *Morphological and acoustical analysis of the nasal and the paranasal cavities.* **J. Acoust. Soc. Am. 96**, 2088-2100.

12. Delattre, P. (1954) *Les attributs acoustiques de la nasalité vocalique et consonantique.* **Studia Linguistica 8**, 103-109.

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

14. Fant, G. (1980) *The relations between area functions and the acoustic signal.* **Phonetica 57**, 55-86.

15. Hawkins, S. and K.N. Stevens (1983) *A cross-language study of the perception of nasal vowels.* **J. Acoust. Soc. Am. 73**, Suppl. 1, S54.

16. Hawkins, S. and K.N. Stevens (1985) *Acoustic and perceptual correlates of the nasal-nonnasal distinction for vowels.* **J. Acoust. Soc. Am. 77**, 1560-1575.

17. Hirano, M., J. Ohala and W. Vennard (1969) *The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation.* **J. Speech Hearing Research 12**, 616-628.

18. Hirsh, I.J. (1959) *Auditory perception of temporal order.* **J. Acoust. Soc. Am. 31**, 759-767.

19. Ishizaka, K., M. Matsudaira, and T. Kaneko (1976) *Input acoustic-impedance measurement of the subglottal system.* **J. Acoust. Soc. Am. 60**, 190-197.

20. Kewley-Port, D. (1982) *Measurement of formant transitions in naturally produced stop consonant-vowel syllables.* **J. Acoust. Soc. Am. 72**, 379-389.

21. Klatt, D.H. and K.N. Stevens (1969) *Pharyngeal consonants.* **RLE Quarterly Progress Report No. 93**, Massachusetts Institute of Technology, Cambridge MA, 207-215.

22. Klatt, D. and L. Klatt (1990) *Analysis, synthesis, and perception of voice quality variations among female and male talkers.* **J. Acoust. Soc. Am. 87**, 820-857.

23. Ladefoged, P. and A. Traill (1994) *Clicks and their accompaniments.* **J. Phonetics 22**, 33-64.

24. Maeda, S. (1982a) *The role of sinus cavities in the production of nasal vowels.* **Proc. of ICASSP-82,** Paris, 911-914.

25. Maeda, S. (1982b) *Acoustic cues of vowel nasalization: A simulation study.* **J. Acoust. Soc. Am. 72**, Suppl. 1, S102.

26. Maeda, S. (1987) *On generation of sound in stop consonants.* **Speech Communication Group Working Papers**, Research Laboratory of Electronics, Massachusetts Institute of Technology, Vol. 5, 1-14.

27. Massey, N.S. (1994) *Transients at stop consonant releases.* S.M. Thesis, Massachusetts Institute of Technology, Cambridge MA.

28. Mrayati, M., R. Carré, and B. Guerin (1988) *Distinctive regions and modes: A new theory of speech production.* **Speech Communication 7**, 257-286.

29. Ohde, R.N. and K.N. Stevens (1983) *Effect of burst amplitude on the perception of stop consonant place of articulation.* **J. Acoust. Soc. Am. 74**, 706-714.

30. Ohman, S.E.G. (1966) *Coarticulation in VCV utterances: Spectrographic measurements.* **J. Acoust. Soc. Am. 39**, 151-168.

31. Pastel, L. (1987) *Turbulent noise sources in vocal tract models.* S.M Thesis, Massachusetts Institute of Technology, Cambridge MA.

32. Peterson, G.E. and H.L. Barney (1952) *Control methods used in a study of the vowels.* **J. Acoust. Soc. Am. 24**, 175-184.

33. Pisoni, D.B. (1977) *Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops.* **J. Acoust. Soc. Am. 61**, 1352-1361.

34. Rothenberg, M. (1968) *The breath-stream of dynamics of simple-released-plostive production.* **Bibliotheca Phonetica, No. 6**, Basel: S. Karger.

35. Rothenberg, M. (1981) *Acoustic interaction between the glottal source and the vocal tract.* In K.N. Stevens and M. Hirano (eds.) **Vocal fold physiology**. Tokyo: University of Tokyo Press, 305-323.

36. Schroeder, M.R. (1967) *Determination of the geometry of the human vocal tract by acoustic measurements.* **J. Acoust. Soc. Am. 41**, 1002-1010.

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

38. Stevens, K.N., A.S. House, and A.P. Paul (1966) *Acoustic description of syllabic nuclei: An integration in terms of a dynamic model of articulation.* **J. Acoust. Soc. Am. 40**, 123-132.

39. Stevens, K.N. (1971) *Airflow and turbulence noise for fricative and stop consonants.* **J. Acoust. Soc. Am. 50**, 1180-1192.

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

41. Stevens, K.N. (1985a) *Spectral prominences and phonetic distinctions in language.* **Speech Communication 4**, 137-144.

42. Stevens, K.N. (1985b) *Evidence for the role of acoustic boundaries in the perception of speech sounds.* In V. Fromkin (ed.) **Phonetic Linguistics**, New York: Academic Press, 243-255.

43. Stevens, K.N. (1989) *On the quantal nature of speech.* **J. Phonetics 17**, 3-46.

44. Stevens, K.N. (1993) *Models for the production and acoustics of stop consonants.* **Speech Communication 13**, 367-375.

45. Stevens, K.N. (in preparation) **Acoustic phonetics.**

46. Syrdal, A.K. and H.S. Gopal (1986) *A perceptual model of vowel recognition based on the auditory representation of American English vowels.* **J. Acoust. Soc. Am. 79**, 1086-1100.

47. Sussman, H.M., H.A. McCaffrey and S.A. Matthews (1991) *An investigation of locus equations as a source of relative invariance for stop place categorization.* **J. Acoust. Soc. Am. 90**, 1309-1325.

48. Titze, I.R. (1992) *Phonation threshold pressure: A missing link in glottal aerodynamics.* **J. Acoust. Soc. Am. 91**, 2926-2935.

49. Traunmüller, H. (1981) *Perceptual dimension of openness in vowels.* **J. Acoust. Soc. Am. 69**, 1465-1475.

50. van den Berg, J.W., J.T. Zantema and P. Doornenbal, Jr. (1957) *On the air resistance and the Bernoulli effect of the human larynx.* **J. Acoust. Soc. Am. 29**, 626-631.

51. Westbury, J.R. (1979) *Aspects of the temporal control of voicing in consonant clusters in English.* **Texas Linguistic Forum 14**, Department of Linguistics, University of Texas, Austin TX.

FIGURE LEGENDS

Fig. 1 (a) Schematized coronal sections showing the configuration of the vocal folds at various instants of time during a cycle of vibration (after Baer, 1975). Successive times are labeled 1-5, with the glottis being closed at 1 and 2, and open at 3, 4, and 5. The cycle is repeated after 5. The arrows in panel 1 show the outward force caused by the subglottal pressure. The vertical lines depict the midline of the glottis. The horizontal and vertical lines above panel 3 indicate dimensions of 1 mm. (b) Schematized representation of the glottal airflow during a cycle of vibration. The points on the waveform correspond roughly to the times in the various panels in (a). The values of volume velocity and time are appropriate for female vocal folds with a subglottal pressure of about 8 cm H_2O.

Fig. 2 Waveform (a) and spectrum (b) of glottal airflow typical of modal voicing for a female voice.

Fig. 3 Spectrum of aspiration noise source. This spectrum is estimated from measurements with models (Shadle, 1985; Pastel, 1987). The source is modeled as an equivalent volume-velocity source to permit comparison with the periodic glottal source.

Fig. 4 Calculated spectra and relative amplitudes of periodic volume-velocity source and turbulence-noise source for two different glottal configurations: a modal configuration in which the glottis is closed over one-half of the cycle (solid lines), and a configuration in which the minimum glottal opening is 0.1 cm^2 (dashed lines). The spectrum for the periodic component gives the amplitude of the individual harmonics. The noise spectrum is the spectrum amplitude in 50-Hz bands. The calculations are based on theoretical models of glottal vibration and of turbulence noise generation (Shadle, 1985; Stevens, 1993).

Fig. 5 Waveform (a) and spectrum (b) of glottal airflow typical of breathy voicing for a female voice. The spectrum shows noise components at high frequency, as indicated by the dashed line.

Fig. 6 Spectrum of sound-pressure source p_s for a mechanical configuration in which air flows through a narrow constriction in a tube and impinges on an obstacle 3 cm downstream from the constriction, for two values of airflow. The diameter of the (circular) constriction is 0.32 cm; the cross-sectional area of the tube is 5.0 cm^2. The sound-pressure source p_s is normalized by dividing by the characteristic impedance $\rho c/5$. Spectrum of p_s is in 300-Hz bands of frequency. For 0 dB on the ordinate, $\frac{5\rho_s}{\rho c} = 1$ cm^3/s. When the area A_t of the tube downstream from the constriction is different from 5 cm^2, the curves should be scaled up by 20 $\log \frac{A_t}{5}$ dB. Curves are based on experimental data of Shadle (1985). (From Stevens, forthcoming.)

Fig. 7 Transfer function for a vocal tract with uniform cross-sectional area and length 15 cm (omitting the effect of the radiation impedance).

Fig. 8 Left: Distribution of sound-pressure amplitude $|p(x)|$ and volume velocity amplitude $|U(x)|$ in a uniform tube (shown at the top) for the first three natural frequencies $F1$, $F2$, and $F3$. Tube is closed at left-hand end and open at right-hand end.

Right: Curves showing the relative magnitude and direction of the shift ΔFn in formant frequency Fn for a uniform tube when the cross-sectional area is decreased at some point along the length of the tube. The abscissa represents the point at which the area perturbation is made. The – sign represents a decrease in formant frequency and the + sign an increase.

Fig. 9 Plot of $F2$ versus $F1$ showing how formants shift when the shape of an acoustic tube is perturbed in different ways. The mid-point represents equally spaced formants for a uniform tube of length 15.4 cm. The lines with arrows indicate how the formant frequencies change when the tube is modified as

shown by the diagrams. The corners of the diagram are labeled with vowel symbols corresponding roughly to the tube shapes. Dimensions are selected to approximate the vocal-tract size of an adult female speaker.

Fig. 10 Average values of $F1$, $F2$, and $F3$ for American English vowels for adult female speakers. The left panel plots $F2$ versus $F1$, and the right panel $F2$ versus $F3$. Data from Peterson and Barney (1952).

Fig. 11 Examples of spectra of vowels produced by female speakers, in which one or more prominences due to subglottal resonances are evident. The arrows identify these prominences. The lines above the spectra are obtained by smoothing the spectra.

Fig. 12 Spectrogram of the utterance [əhet], together with spectra sampled (a) during the aspiration noise in [h], (b) near the onset of glottal vibration, and (c) in the vowel. The waveforms are shown below the spectra.

Fig. 13 Spectrograms of the syllables [bɛ], [dɛ] and [gɛ], from left to right. These spectrograms illustrate the different transitions of $F2$ and $F3$ following the release of the different consonants. The first-formant frequency $F1$ rises following the release for all the consonants.

Fig. 14 Illustrating the direction of movements of $F2$ and $F3$ when a consonant produced with a constriction at different places in the oral cavity is followed by a vowel with a uniform vocal-tract shape. The formant frequencies for the vowel are 1500 and 2500 Hz. The area functions for the various vocal-tract configurations for the consonants and the vowel are schematized.

Fig. 15 Showing the approximate trajectories of $F2$ and $F3$ movements in the $F2$-$F3$ plane when consonants with different places of articulation are followed by different vowels. The vowel targets are identified by the vowel symbols. For each consonant (labial, alveolar, or velar) the movement of $F2$ and $F3$ when the consonant is released into the vowel is indicated by a line with an arrow.

The starting points of the $F2$ and $F3$ movements for three classes of following vowels are identified by closed contours: front vowels, unrounded back vowels, and rounded back vowels. The vowel formant frequencies are from Peterson and Barney (1952). The data on formant movements for consonants are taken from various sources, including unpublished measurements. Some smoothing has been applied to the data. These charts apply to adult male speakers. For adult females, the frequencies should be scaled up by 10-20 percent.

Fig. 16 This chart is organized like those in Fig. 15, except that the consonants are the pharyngeal fricative [ħ] and the uvular fricative [χ], and only three vowels are shown. Data are from Alwan (1986), and are averages for several male and female speakers of Arabic.

Fig. 17 The solid line shows a typical spectrum envelope for a nonnasal, nonhigh vowel in the $F1$ region. The dashed line illustrates how this spectrum is modified when the same vowel is nasalized. The overall effect of nasalization is a flattening of the spectrum in the $F1$ region.

Fig. 18 Spectrograms and spectra illustrating the acoustic attributes of nasal and nonnasal vowels in French. The left column shows a spectrogram of the word *combat*, and the two spectra below are sampled in the nasal vowel and in the nonnasal vowel, respectively. Similar displays in the right column are for the word *engage*.

Fig. 19 Illustrating how the spectrum changes at the release of a nasal consonant into a following vowel. At the top is a spectrogram of the utterance *a knock*. In the column below the spectrogram are two spectra: one sampled in the nasal murmur for [n] and the other sampled immediately following the release. The formant frequencies $F2$ and $F3$ are labeled, as is the nasal formant FN. Similar pairs of spectra are given in the left column for the utterance *a mill* and in the right column for the utterance *sing out*. The solid line above each spectrum is a smoothed version of the spectrum. The increase in spectrum

amplitude in the $F2$ region is about 20 dB as the consonant-vowel boundary is crossed.

Fig. 20 The top panels show spectra sampled during the constricted interval for liquids and glides in various vowel environments, as shown. The spectra in the bottom panels are sampled during the initial or middle parts of the following vowels. The waveforms are shown below the spectra, together with the time windows for the spectra. The time intervals from top spectrum to bottom spectrum are as follows: for [lʌ], 36 ms; for [rʌ], 93 ms; for [wu], 97 ms; for [ji], 101 ms. Speaker is adult female.

Fig. 21 The top panel shows the estimated change in cross-sectional area of the consonantal constriction when a voiceless unaspirated alveolar or labial stop consonant is produced in intervocalic position. The calculated intraoral pressure, based on an aerodynamic and mechanical model of the vocal tract, is also displayed. The middle panel gives the calculated airflows during the production of the consonant: the flows U_c through the constriction, U_g through the glottis, and U_w due to outward movement of the vocal-tract walls. The bottom panel shows estimates of the instantaneous flow, and depicts each glottal pulse in the vicinity of the consonant interval. The vertical lines in the bottom two panels show the times of closure and opening. The subglottal pressure is 8 cm H_2O.

Fig. 22 Schematic representation of sequence of events at the release of a voiceless unaspirated stop consonant. A typical waveform (with time scale) is shown at the bottom.

Fig. 23 At the top is a spectrogram of the utterance [ˈbɑpə], illustrating the acoustic attributes of a voiceless unaspirated [p]. The waveform of the portion of the utterance from near the end of the first vowel to the first few glottal periods of the second vowel is shown below, with the consonant release marked by an arrow.

Fig. 24 Each panel shows the spectrum sampled in the burst at the release of a voiceless unaspirated stop consonant (solid line), together with the spectrum near the second glottal period in the vowel (dashed line). The spectra are obtained by averaging a series of spectra (6.4-ms time window) sampled at 1-ms intervals over 10 ms. From top to bottom, the consonants are labial, alveolar, and velar, and the following vowel is [ɑ]. Adult male speaker.

Fig. 25 (a) Schematized time variation of cross-sectional area of vocal-tract constriction A_c and glottal constriction A_g for production of a voiceless fricative consonant in intervocalic position. The solid lines indicate what the constriction areas would be if there were no increase in intraoral pressure, and the dashed lines show how the areas are modified by forces on the structures due to the increased intraoral pressure. (b) Calculated intraoral pressure P_m and transglottal pressure ΔP_g. (c) Calculated airflow.

Fig. 26 Top: Spectrogram of the utterance [ʌsʌ] produced by an adult male speaker. Bottom: The solid line is the measured spectrum for [s]. The dashed line is the spectrum sampled in the vowel near the second glottal period. Spectra are measured as in Fig. 24, except that the averaging time for the spectrum of [s] is 100 ms.

Fig. 27 Spectra of four voiceless fricatives in English, as labeled. The spectra are obtained with a 6.4 ms time window over a time interval of 100 ms. The fricatives were produced in isolated VCV syllables, with the vowel [ʌ]. The spectra are given with approximately the correct relative amplitudes. Adult male speaker.

Fig. 28 Spectrograms and spectra illustrating some of the acoustic attributes of (a) a voiced unaspirated stop consonant (in the utterance *a bill*), (b) a voiceless aspirated stop consonant (in *a pet*), and (c) a voiced aspirated stop consonant (in *dha*). Spectrograms are given at the top. In (a) the top panel shows the spectrum (6.4 ms window) sampled in the middle of the consonant (dashed line)

and in the second glottal pulse after release. The bottom panel is the spectrum of the burst. In (b) the spectra are the burst (top), the aspiration (middle), and immediately after voicing onset (bottom, with a 26-ms window). In (c) the top spectrum is in the aspiration and the bottom spectrum is immediately after voicing onset (26-ms window).

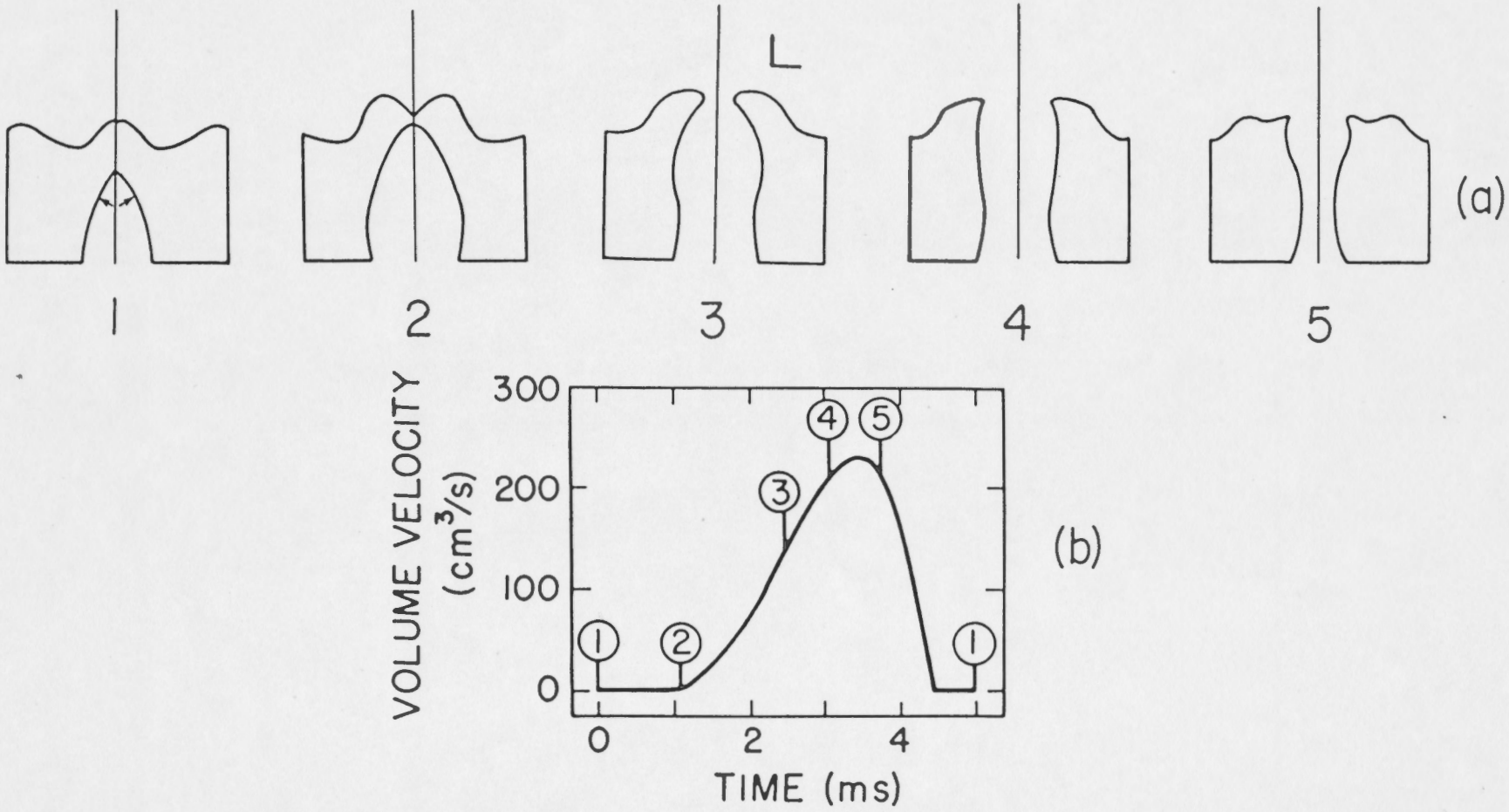


Fig. 1

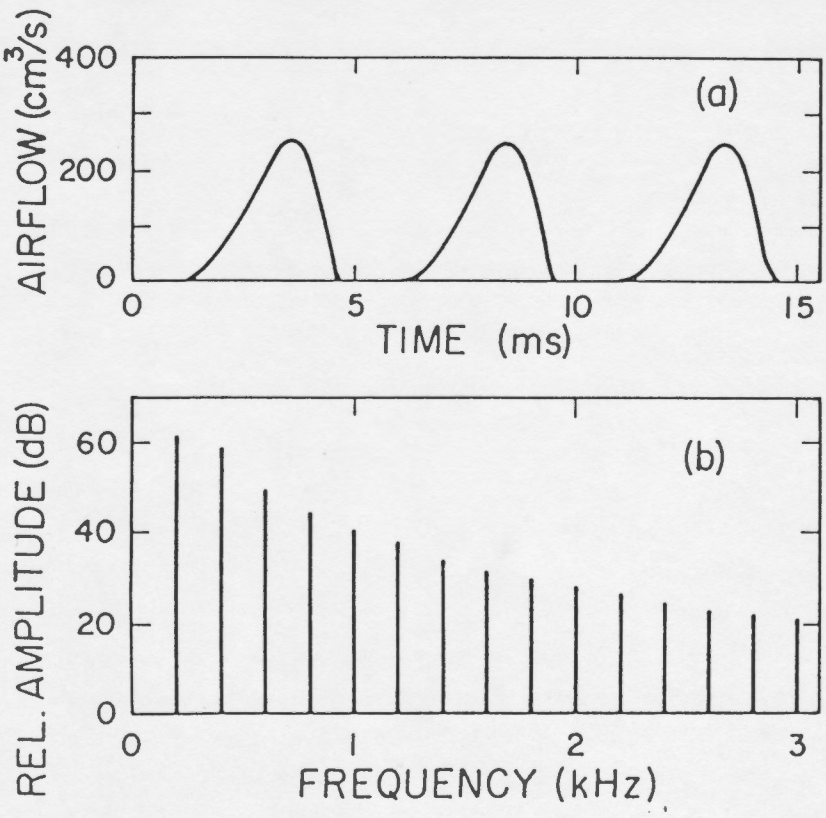

AIRFLOW (cm³/s)
400
200
0
(a)
0
5
10
15
TIME (ms)
REL. AMPLITUDE (dB)
60
40
20
0
(b)
0
1
2
3
FREQUENCY (kHz)

Fig.2

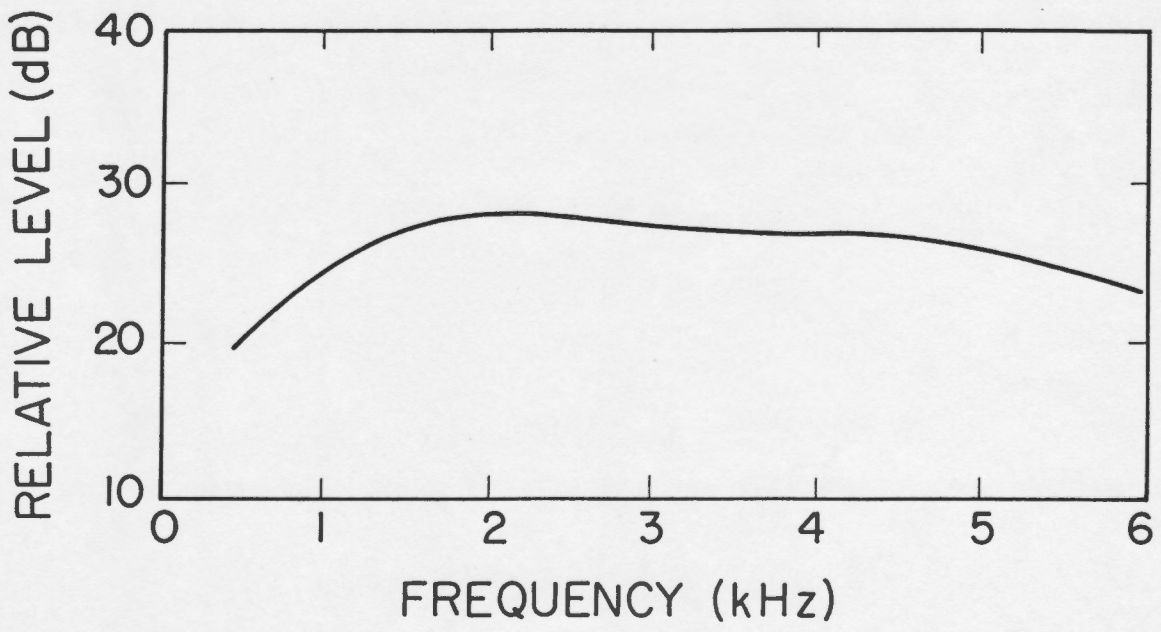


Fig.3

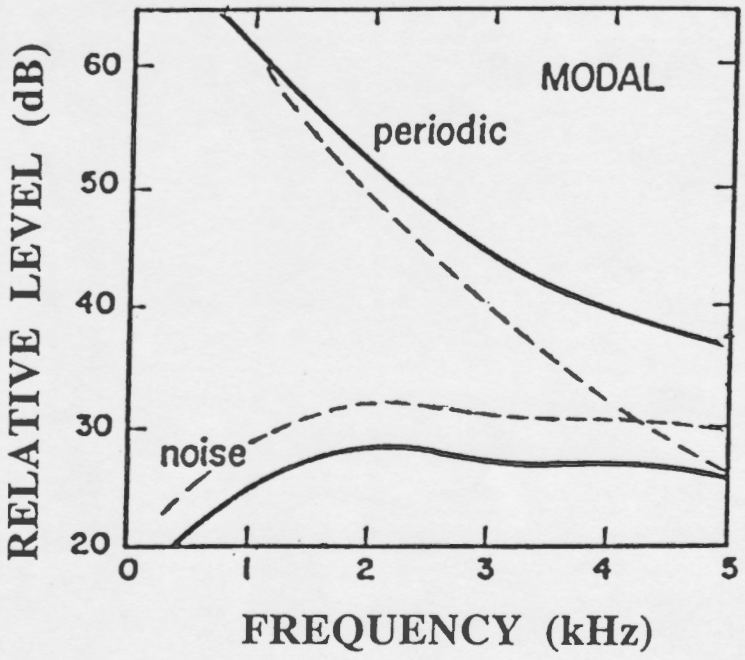

RELATIVE LEVEL (dB)
60
50
40
30
20
0
1
2
3
4
5
FREQUENCY (kHz)
MODAL
periodic
noise

Fig.4

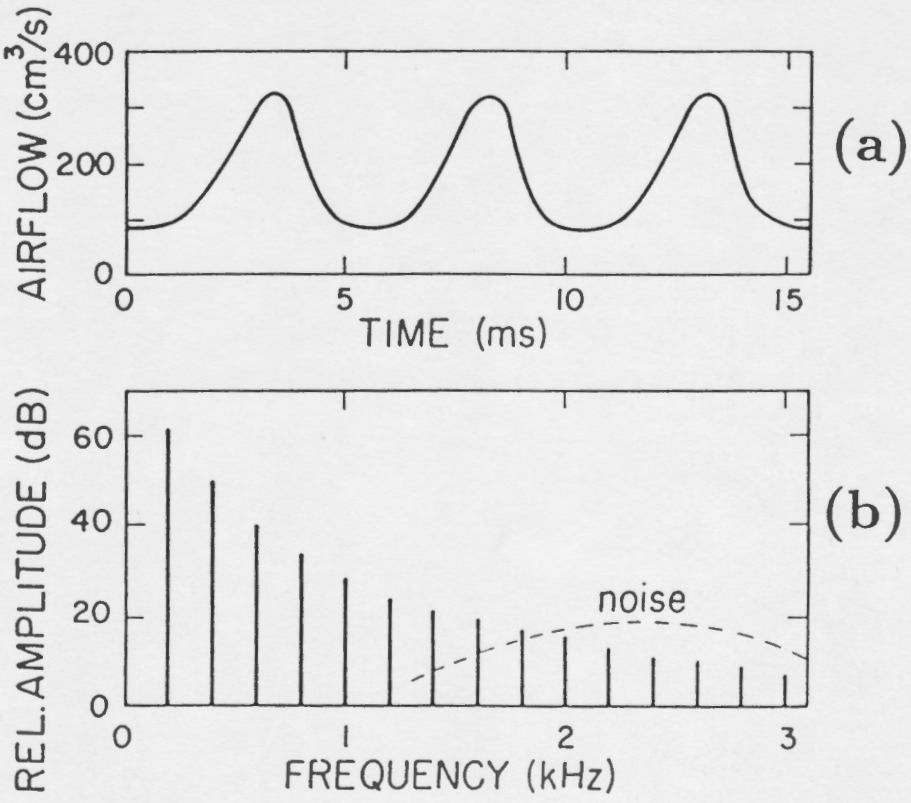

AIRFLOW (cm³/s)
400
200
0
0
5
10
15
TIME (ms)
(a)
REL. AMPLITUDE (dB)
60
40
20
0
noise
0
1
2
3
FREQUENCY (kHz)
(b)

Fig.5

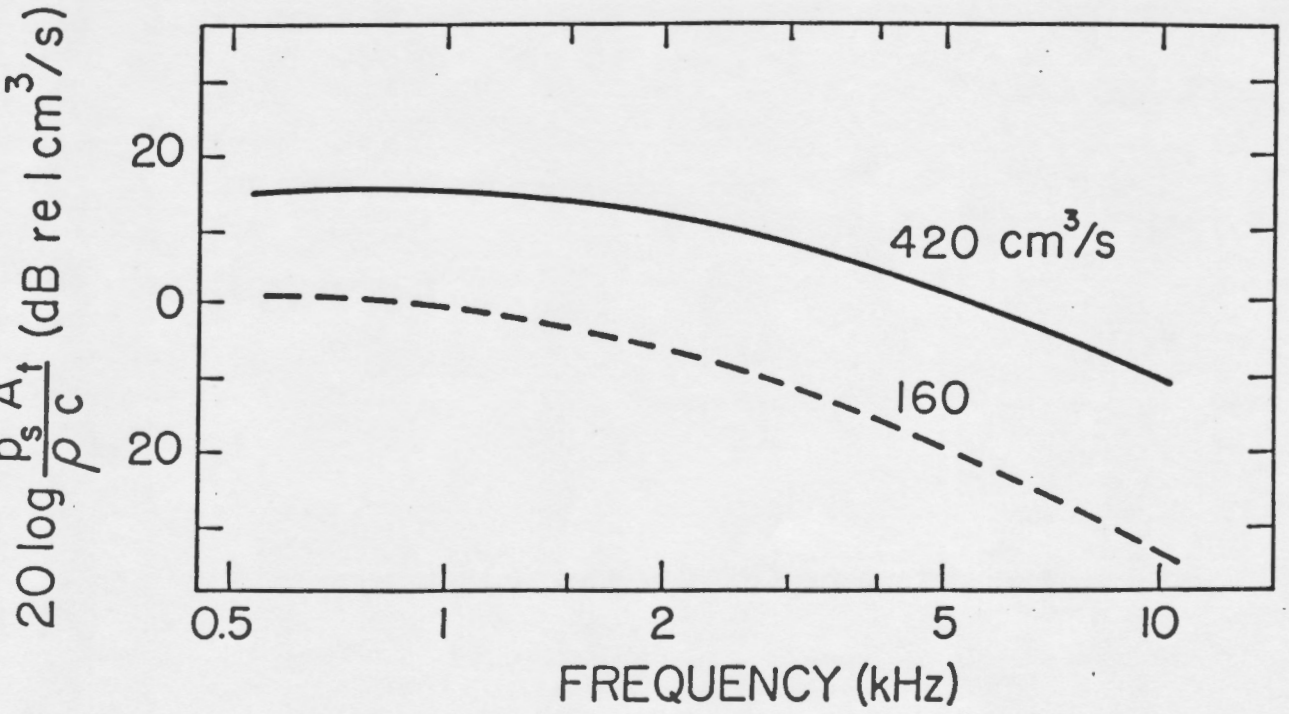

20 log $\frac{p_s A_t}{\rho c}$ (dB re 1 cm^3/s)
20
0
20
420 cm^3/s
160
0.5
1
2
5
10
FREQUENCY (kHz)

Fig. 6

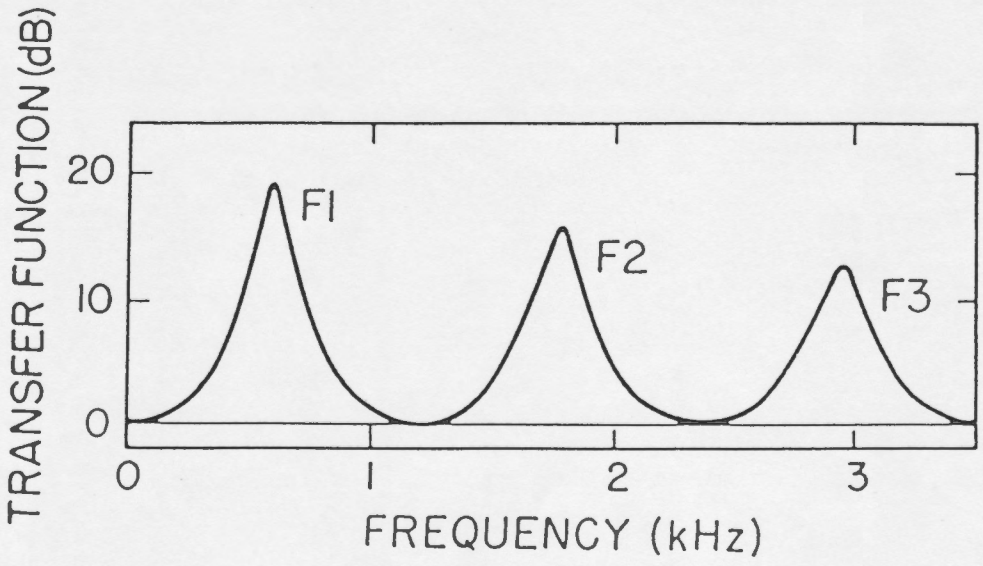


Fig.7

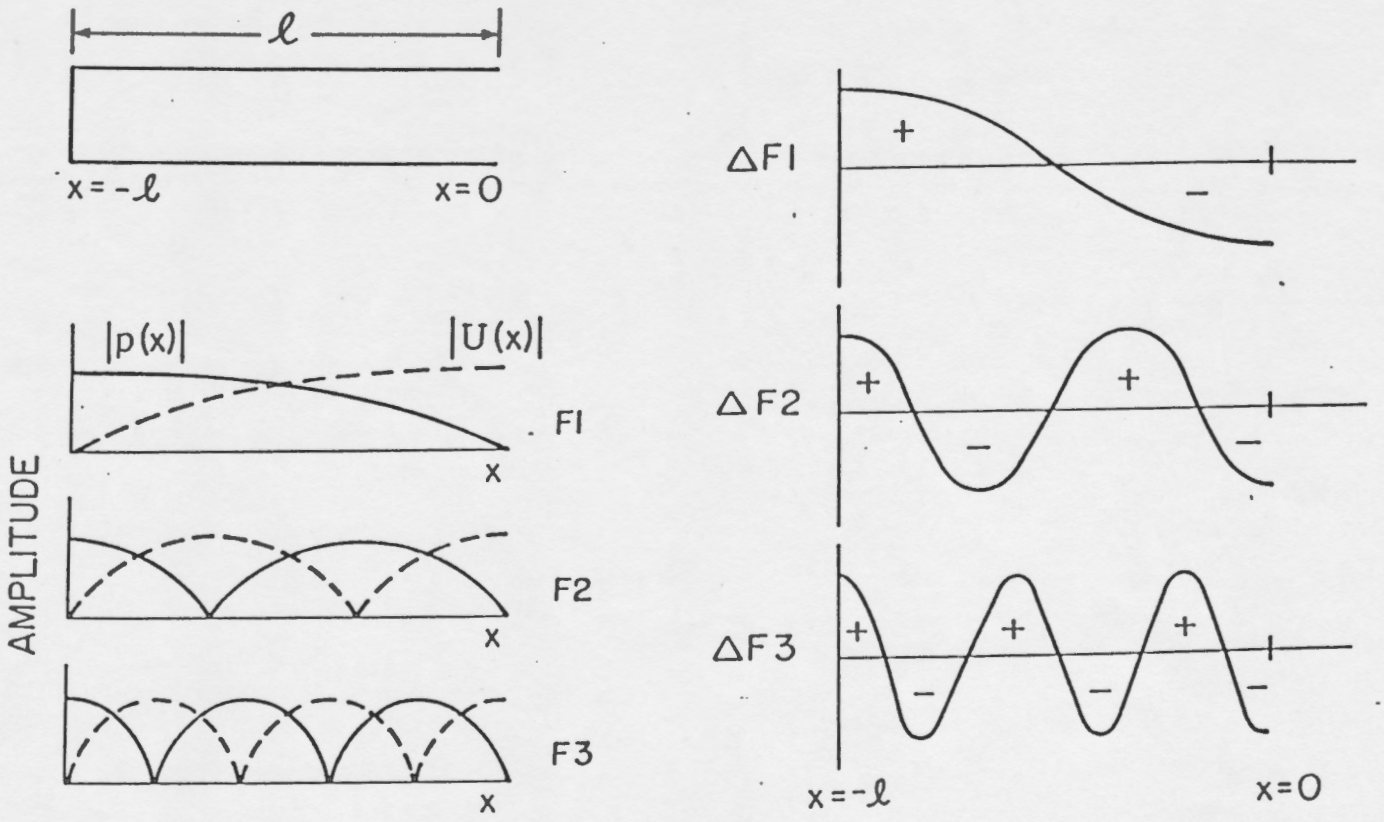


Fig.8

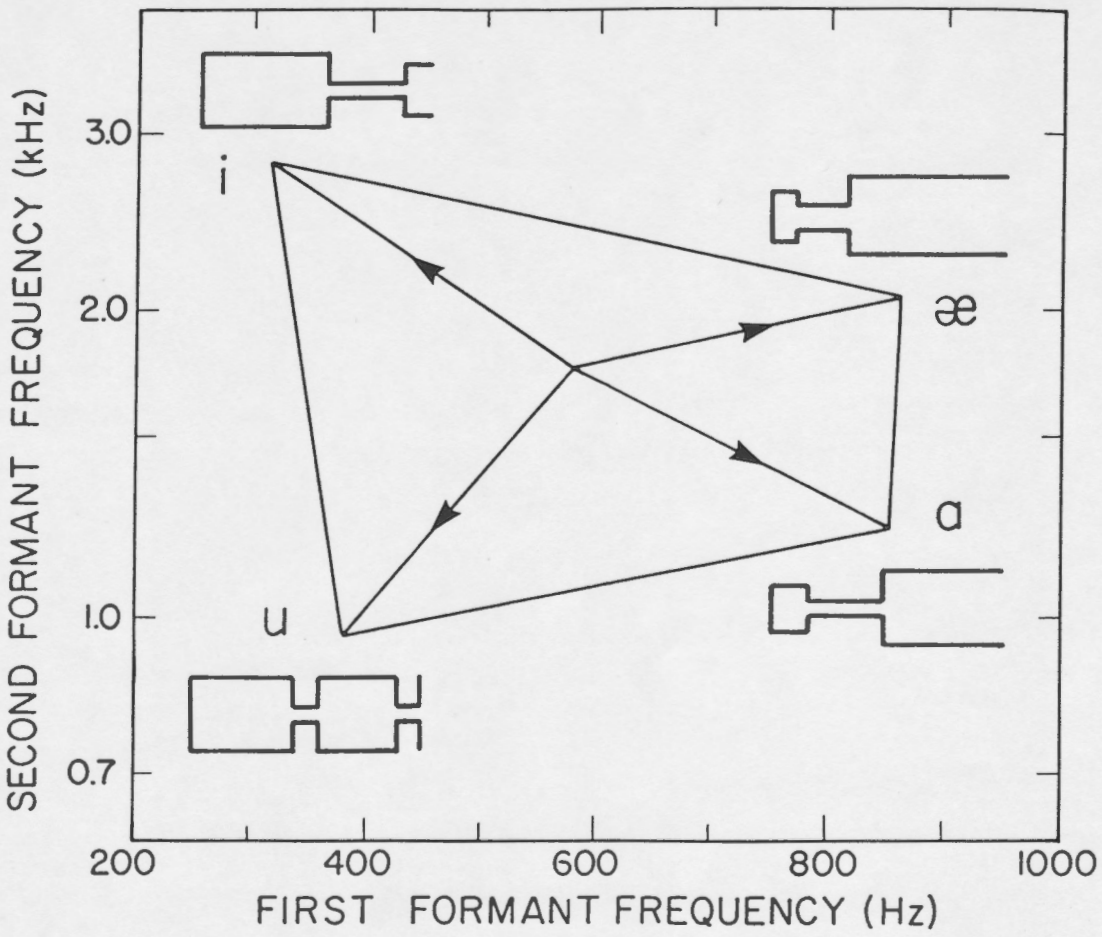

SECOND FORMANT FREQUENCY (kHz)
3.0
2.0
1.0
0.7
i
æ
a
u
200
400
600
800
1000
FIRST FORMANT FREQUENCY (Hz)

Fig.9

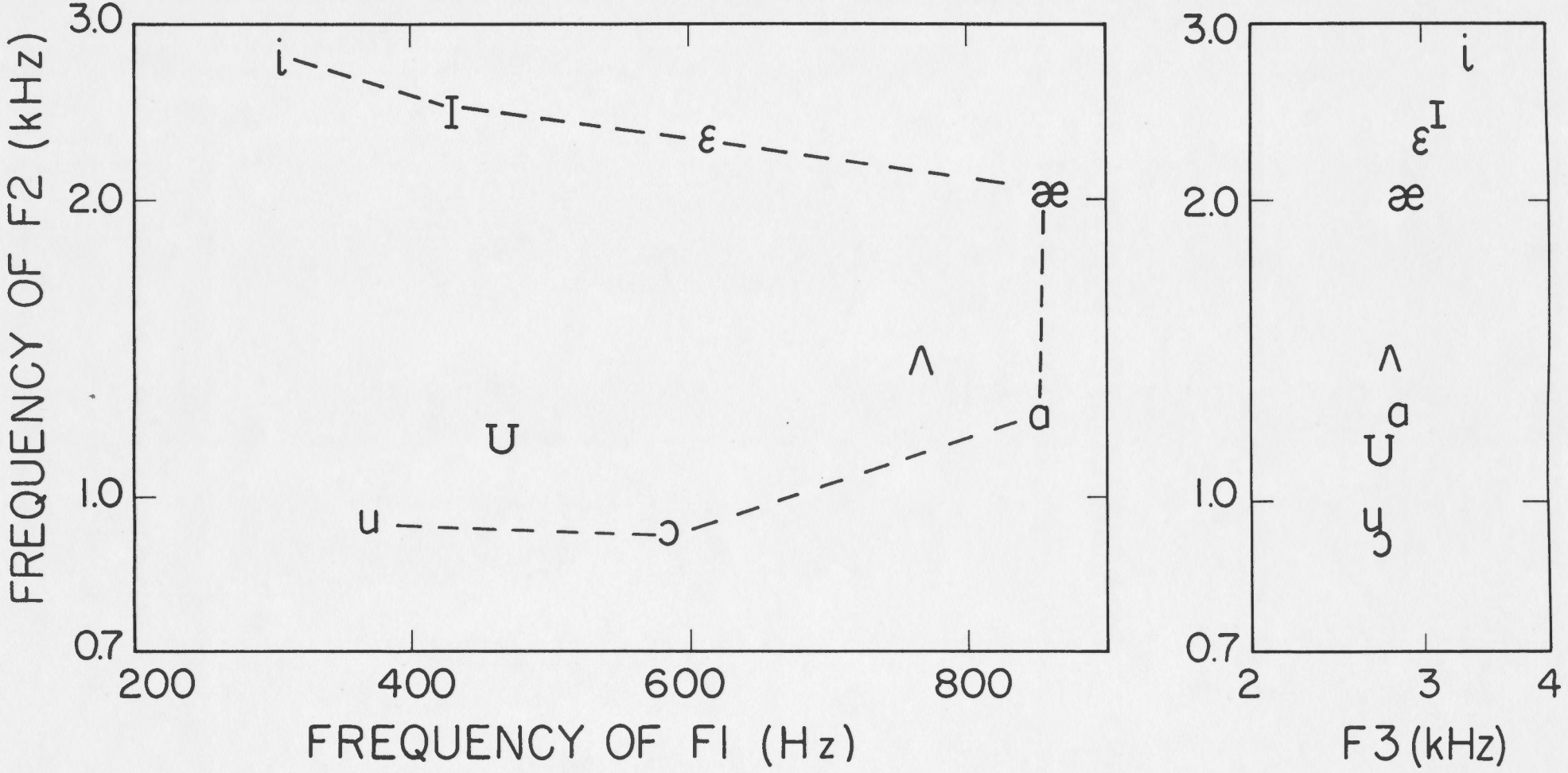

3.0
2.0
1.0
0.7
FREQUENCY OF F2 (kHz)
200
400
600
800
FREQUENCY OF FI (Hz)
i
I
ε
æ
ʌ
a
U
u
ɔ
2
3
4
F3 (kHz)

Fig.10

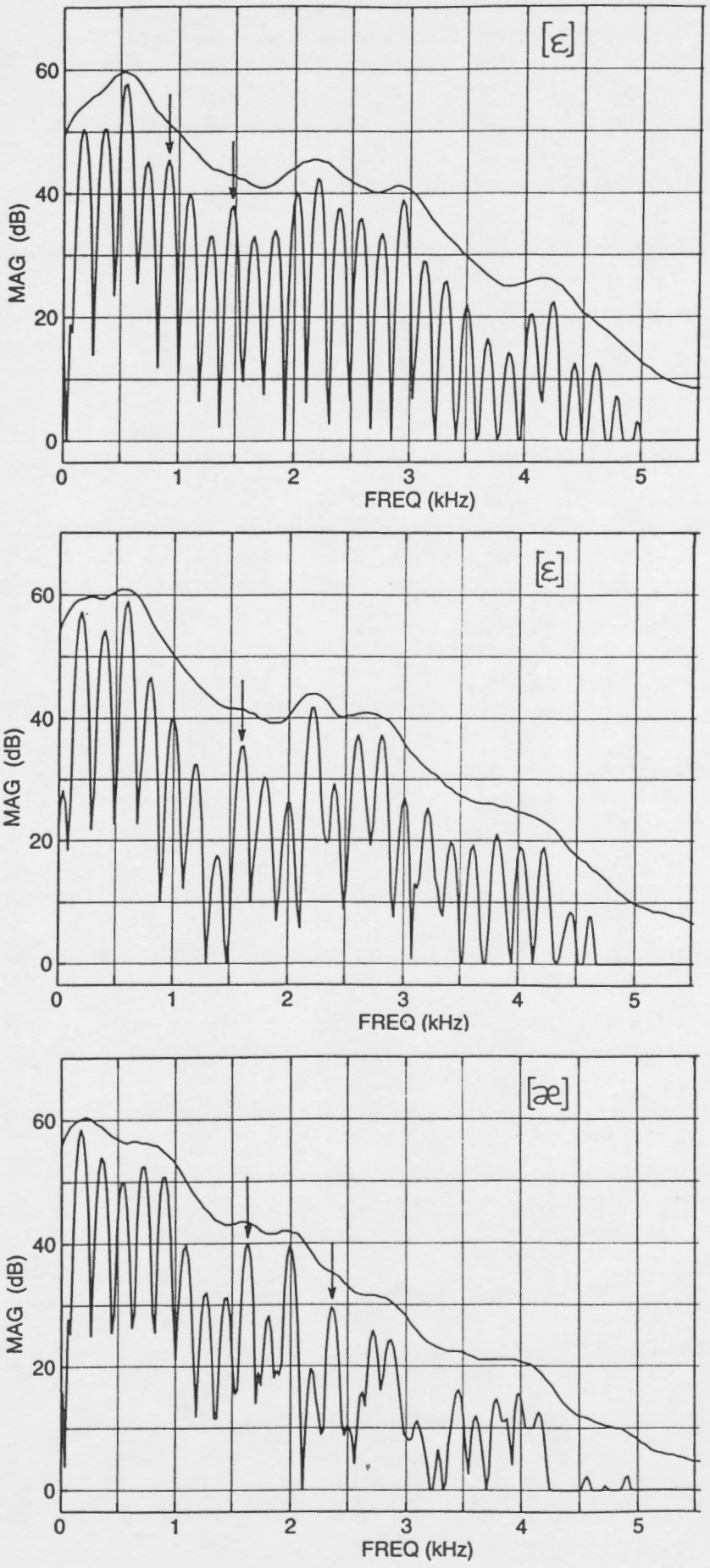

[ɛ]
MAG (dB)
FREQ (kHz)
0
20
40
60
0
1
2
3
4
5
[ɛ]
MAG (dB)
FREQ (kHz)
[æ]
MAG (dB)
FREQ (kHz)

Fig. 11

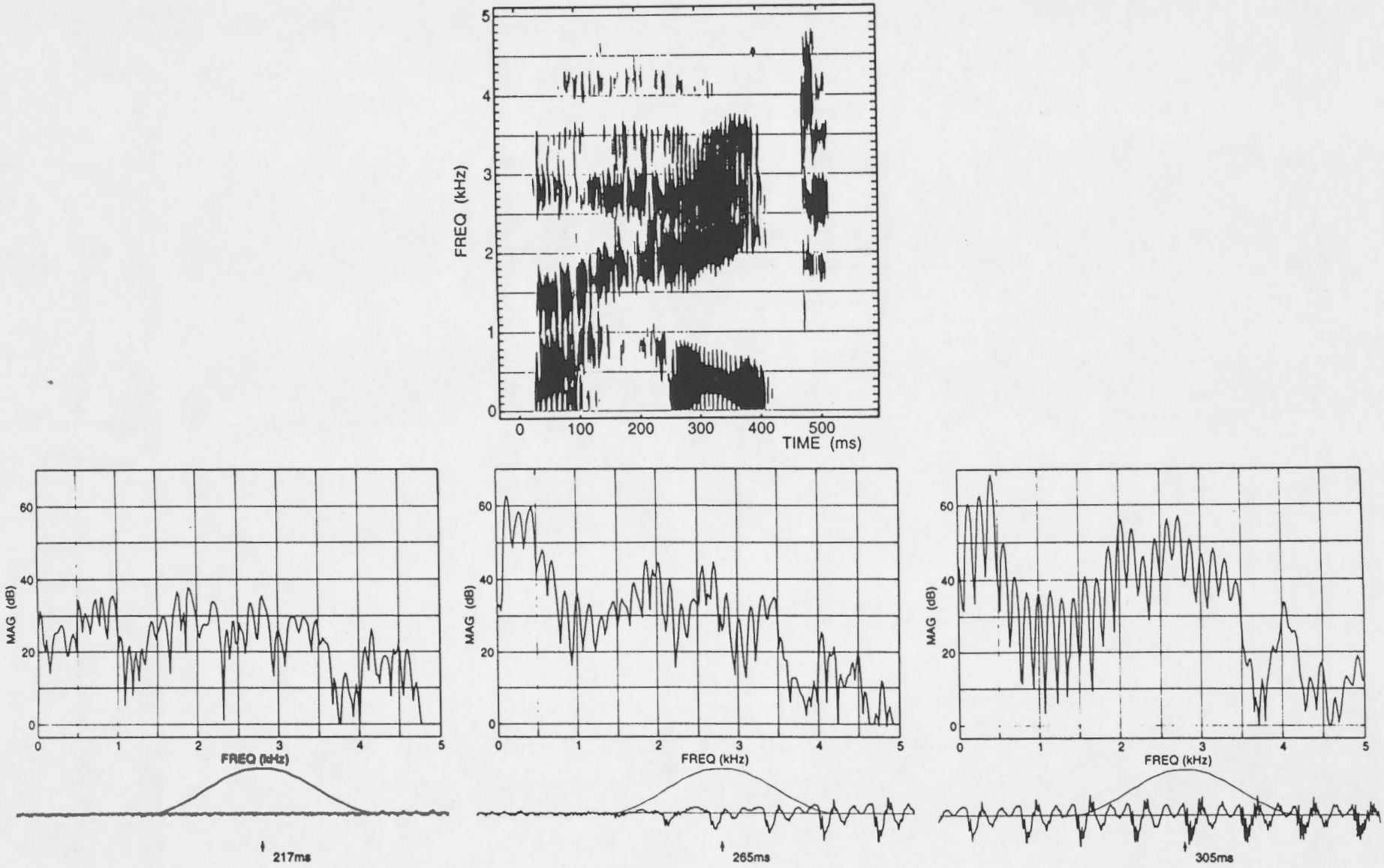


Fig. 12

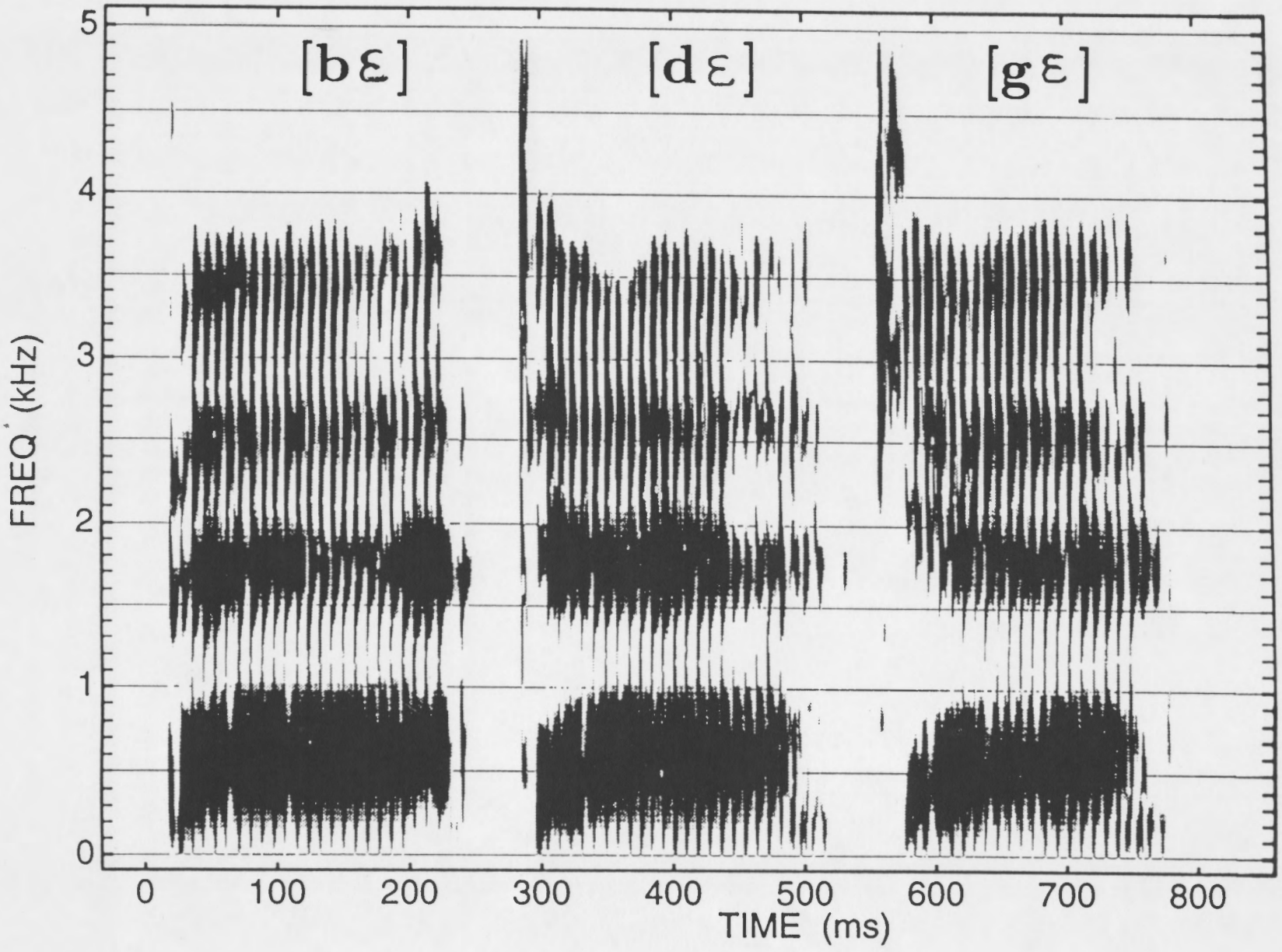

[bɛ]
[dɛ]
[gɛ]
FREQ (kHz)
5
4
3
2
1
0
0
100
200
300
400
500
600
700
800
TIME (ms)

Fig. 13

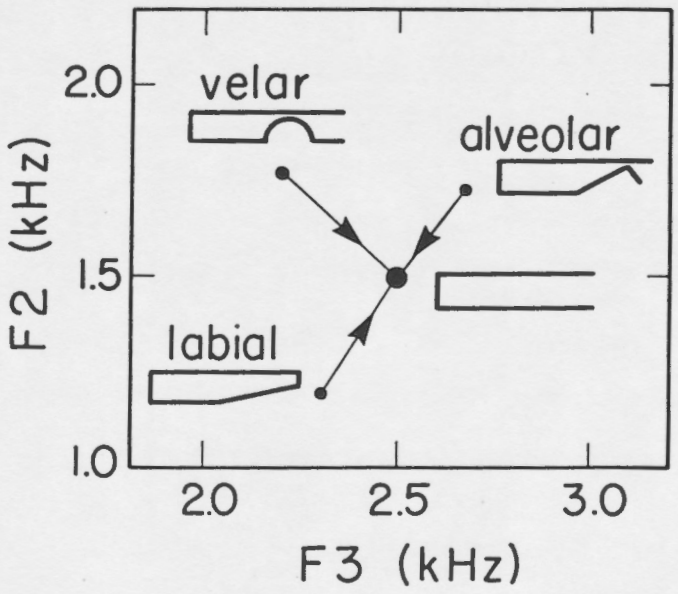


Fig.14

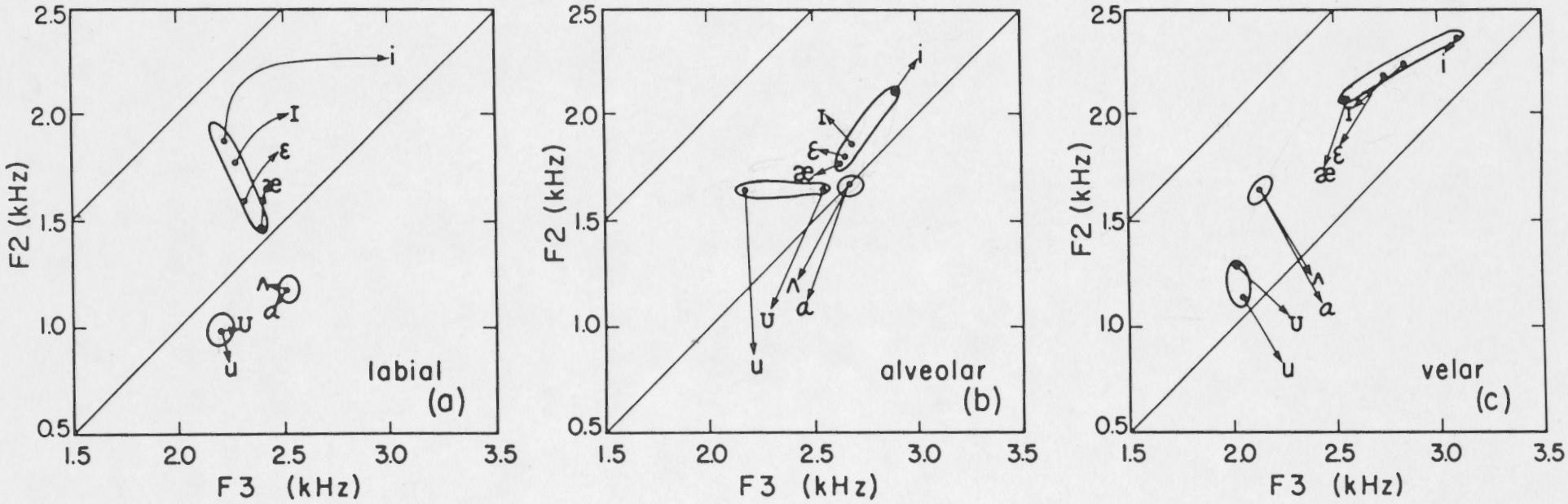

2.5
2.0
1.5
1.0
0.5
F2 (kHz)
1.5
2.0
2.5
3.0
3.5
F3 (kHz)
i
I
ɛ
æ
ʌ
ɑ
ʊ
u
labial
(a)
alveolar
(b)
velar
(c)

Fig. 15

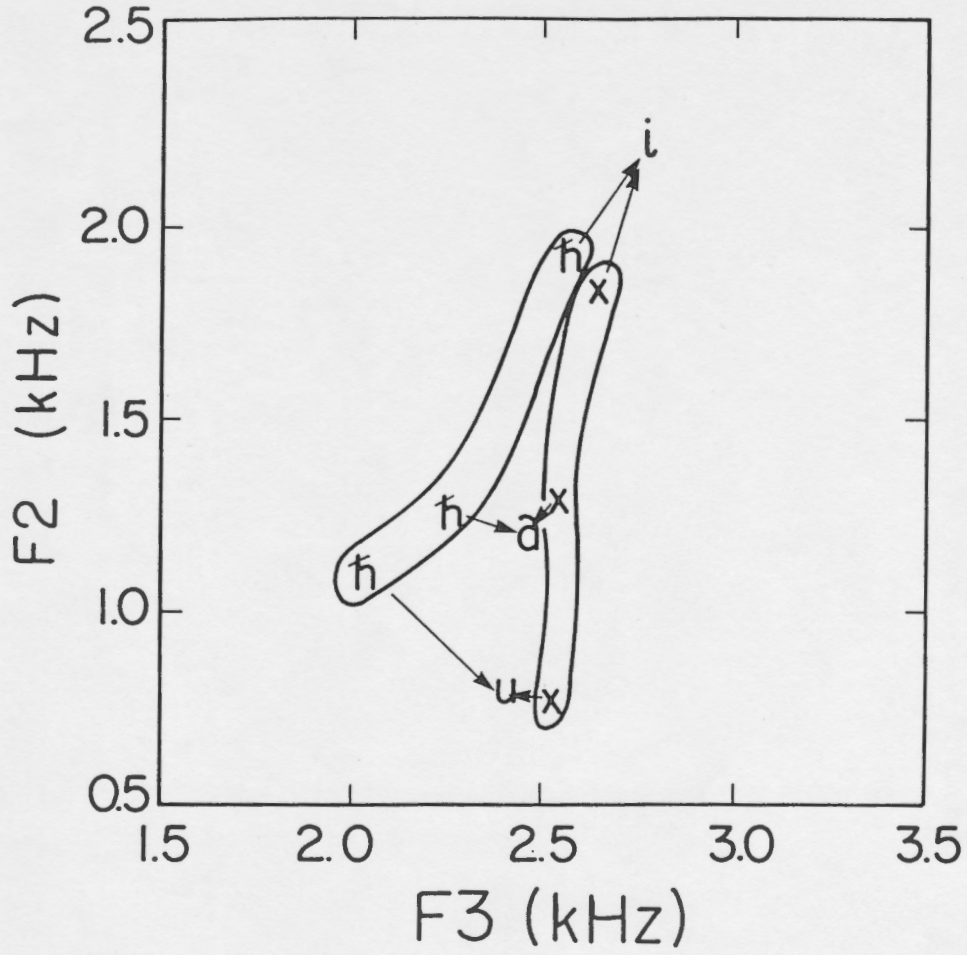


Fig.16

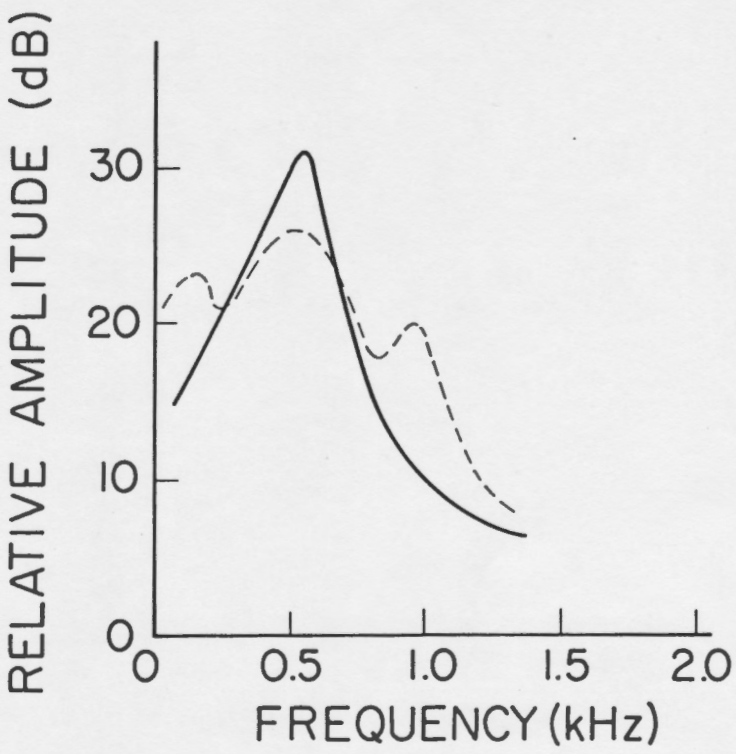


Fig.17

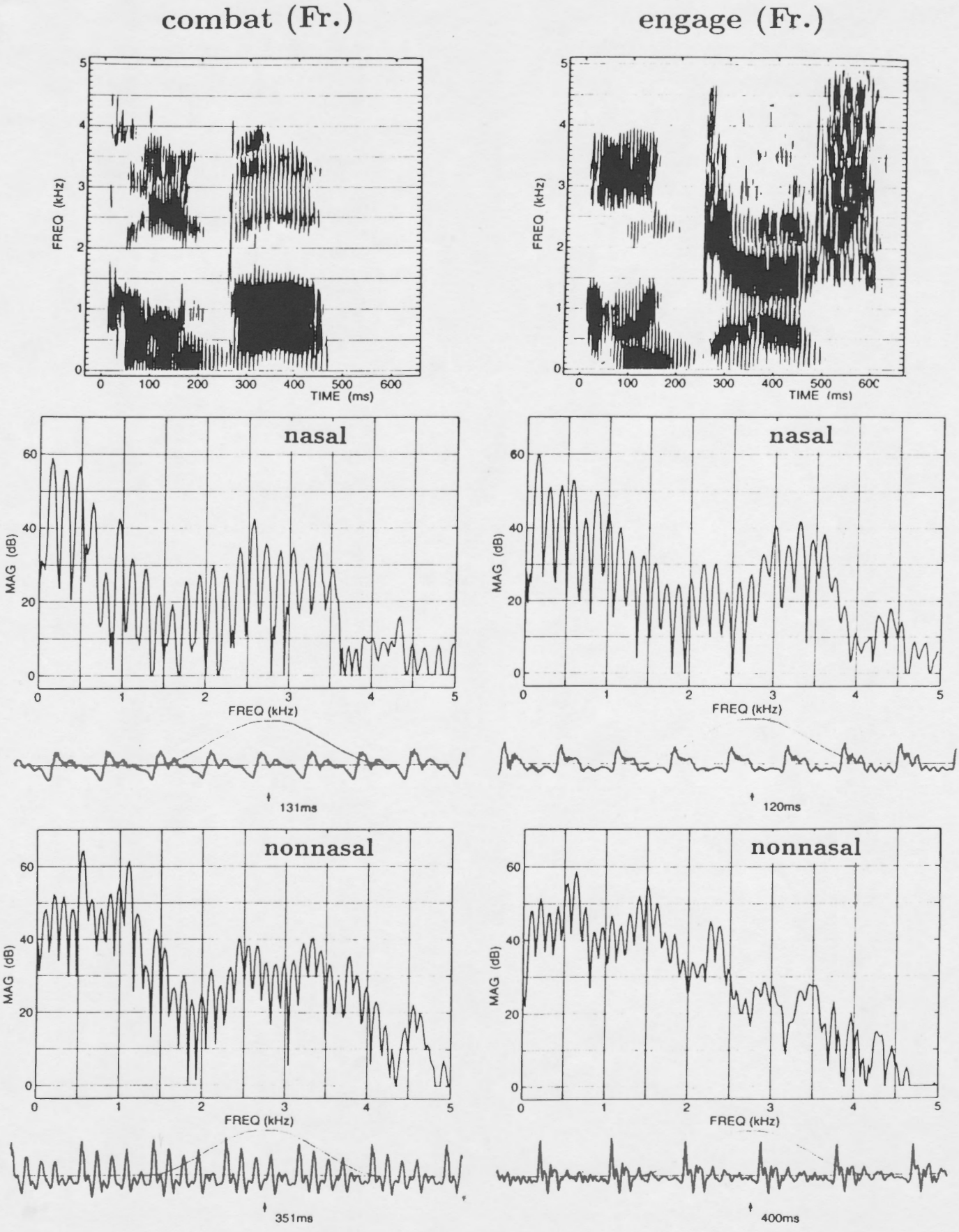

combat (Fr.)
engage (Fr.)
FREQ (kHz)
TIME (ms)
nasal
MAG (dB)
FREQ (kHz)
131ms
120ms
nonnasal
351ms
400ms

Fig. 18

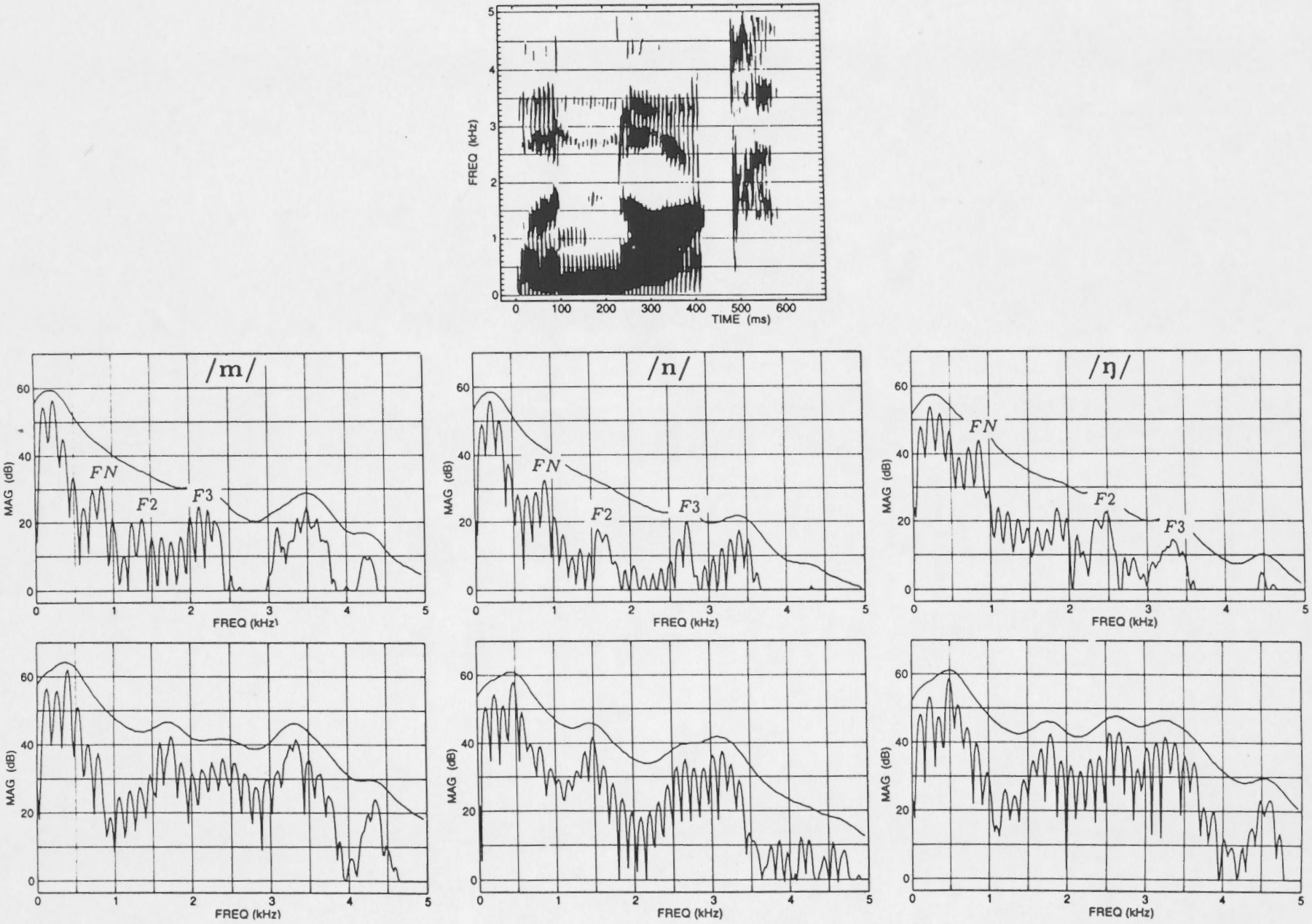


Fig. 19

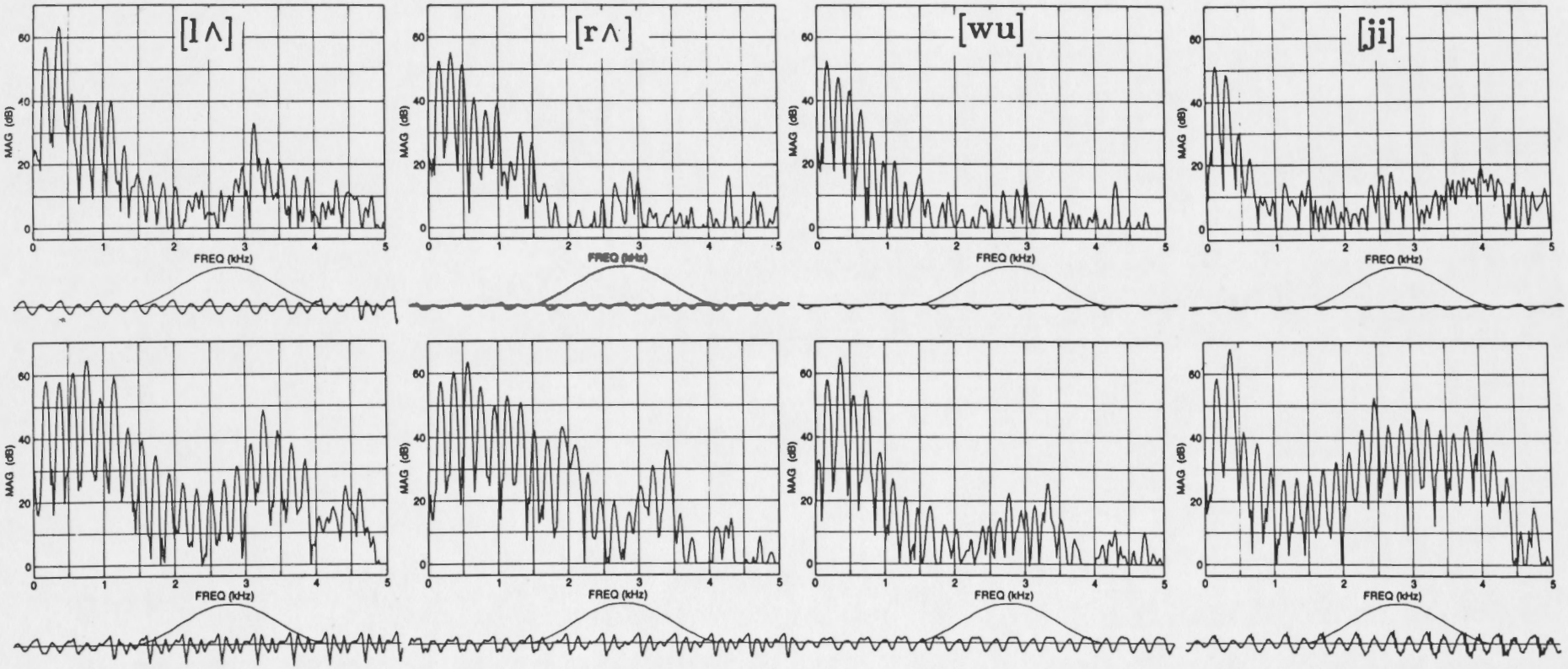


Fig.20

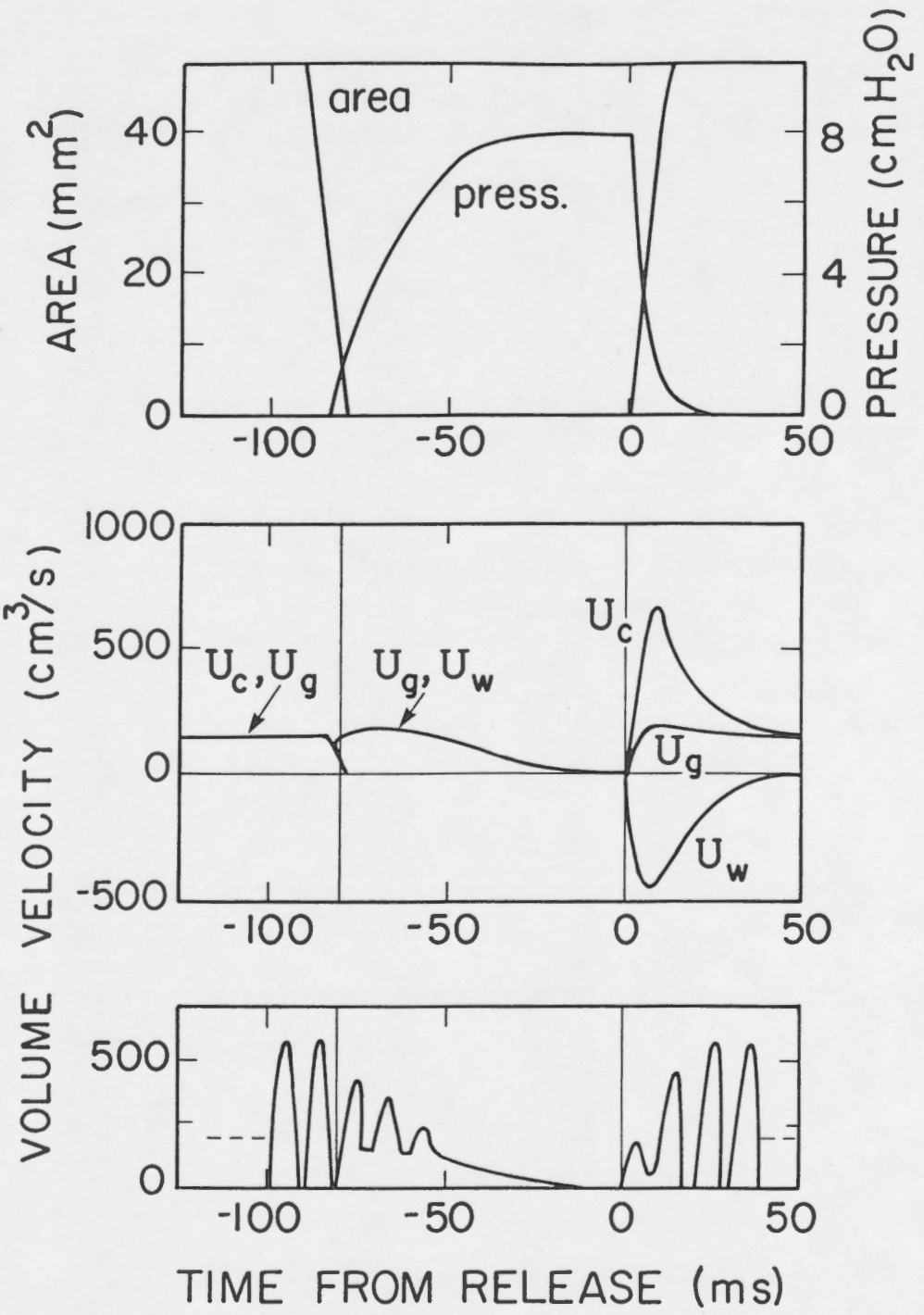

area
press.
AREA (mm²)
PRESSURE (cm H₂O)
40
20
0
8
4
0
-100
-50
0
50
1000
500
0
-500
VOLUME VELOCITY (cm³/s)
Uc, Ug
Ug, Uw
Uc
Ug
Uw
500
0
TIME FROM RELEASE (ms)

Fig.21

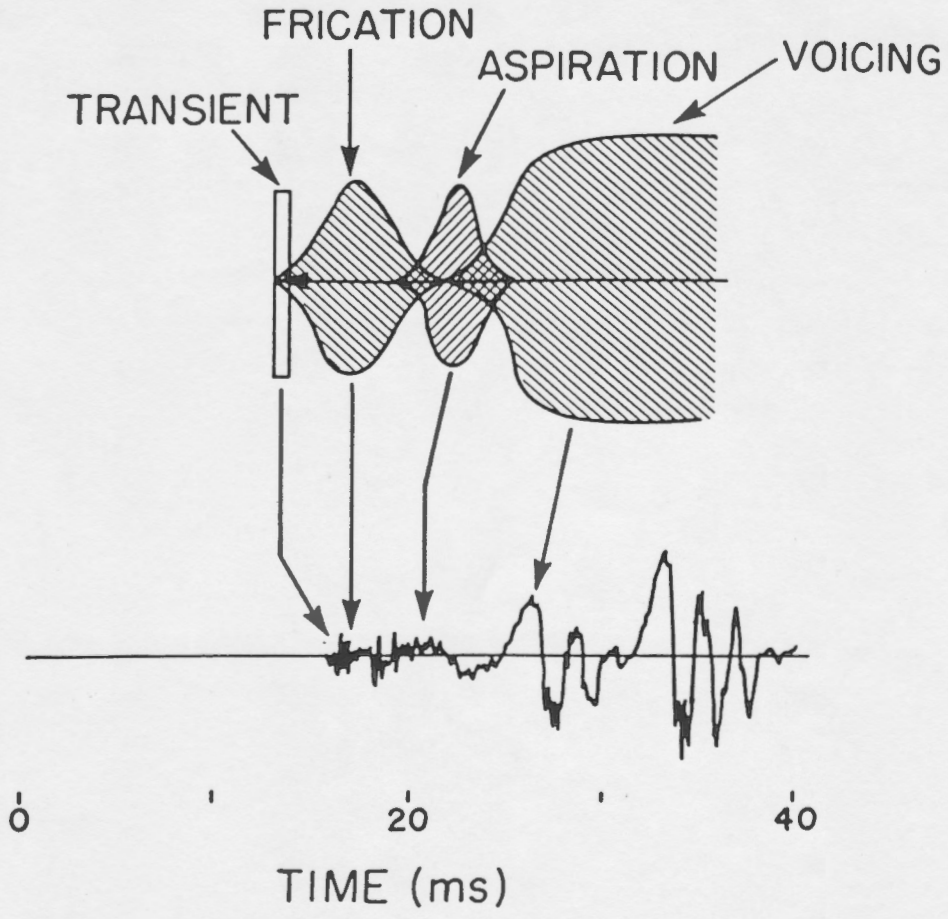


Fig.22

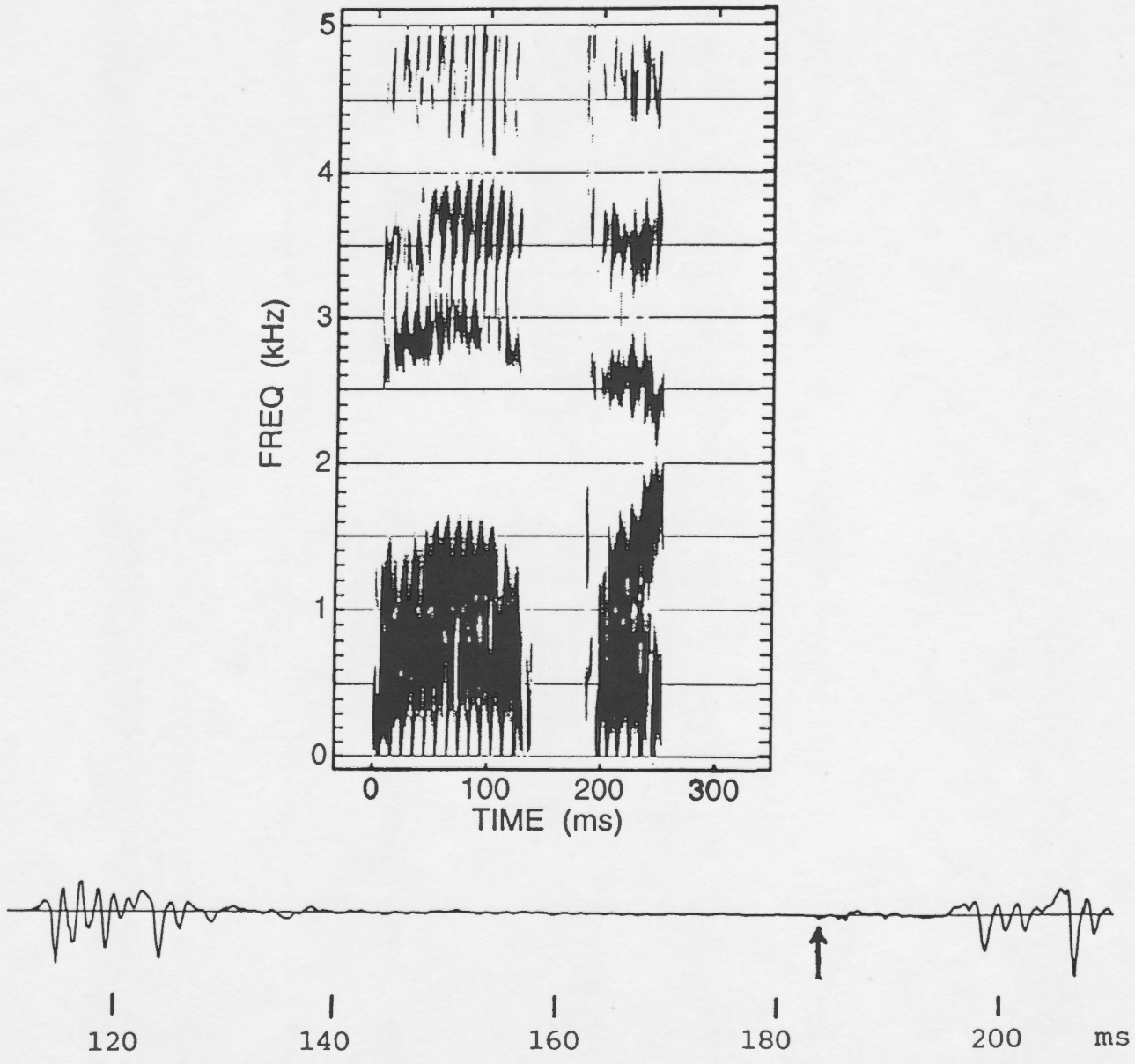
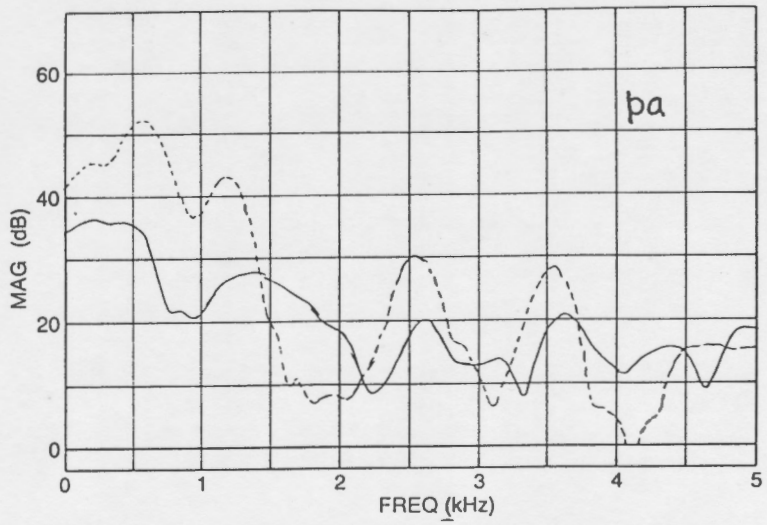
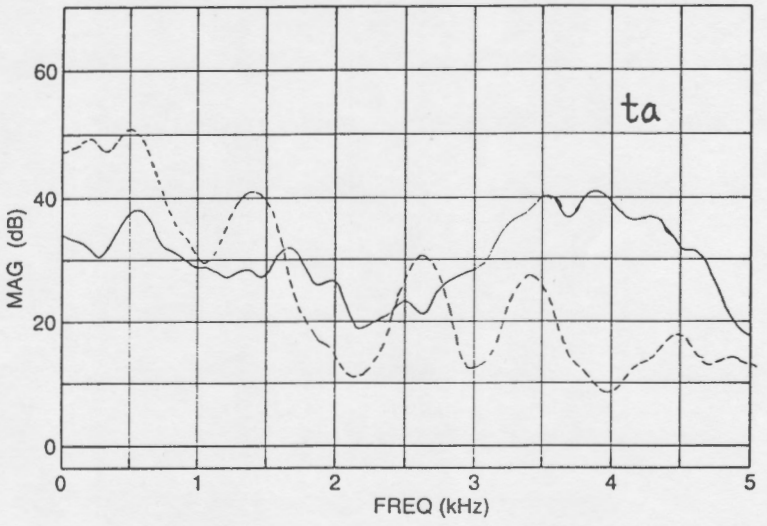


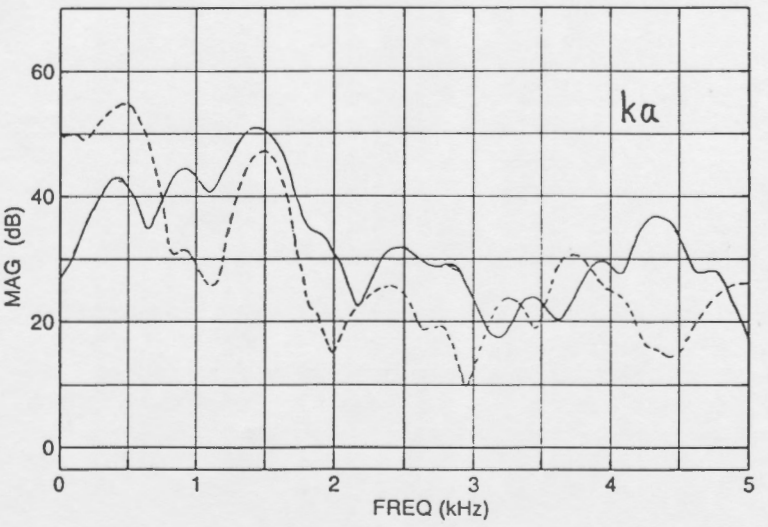
Fig. 23


pa
MAG (dB)
FREQ (kHz)
0
20
40
60
0
1
2
3
4
5

(a)


ta
MAG (dB)
FREQ (kHz)
0
20
40
60
0
1
2
3
4
5

(b)


ka
MAG (dB)
FREQ (kHz)
0
20
40
60
0
1
2
3
4
5

(c)

Fig.24

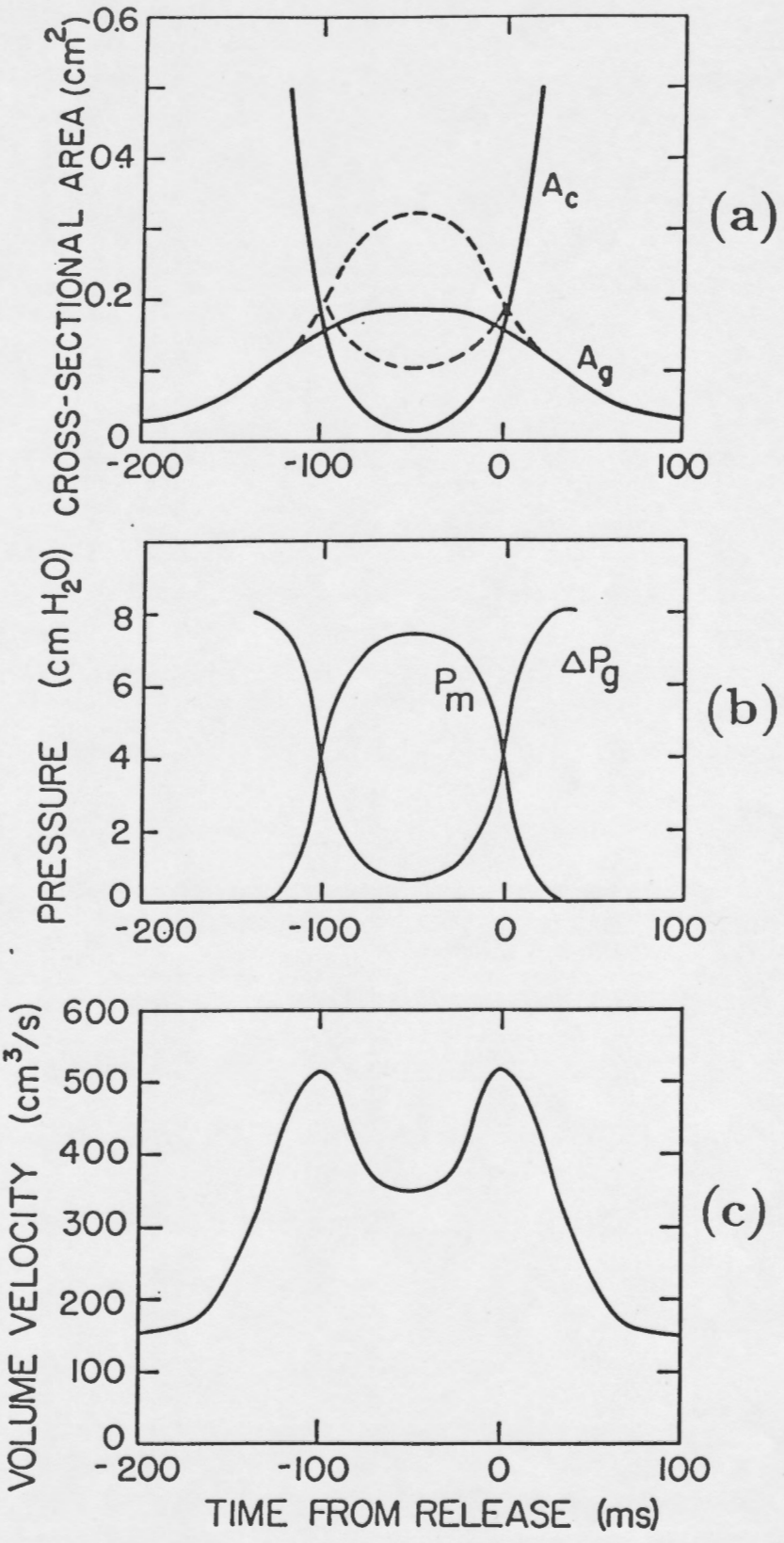
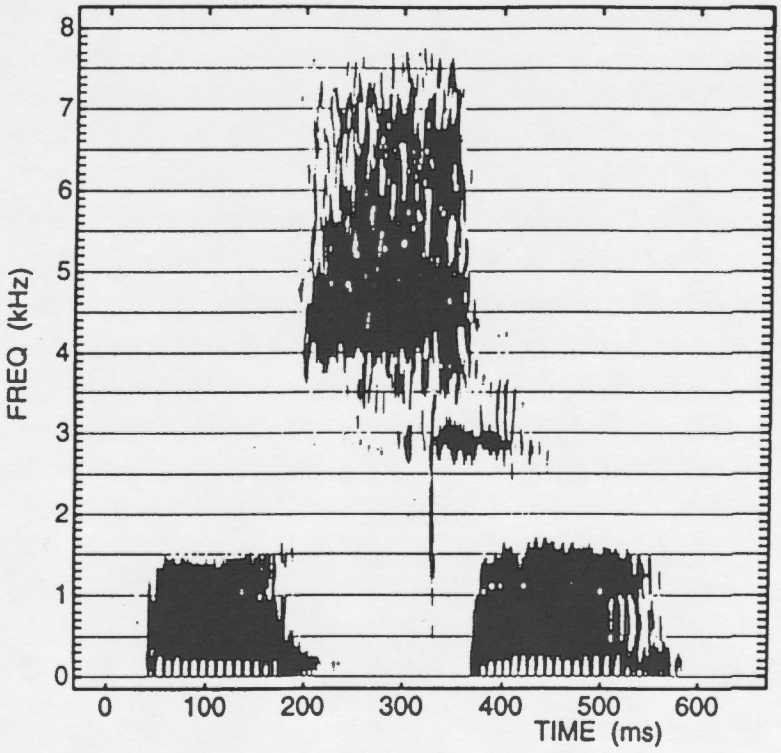

CROSS-SECTIONAL AREA (cm²)
0.6
0.4
0.2
0
A_c
A_g
(a)
-200
-100
0
100
PRESSURE (cm H_2O)
8
6
4
2
0
P_m
ΔP_g
(b)
-200
-100
0
100
VOLUME VELOCITY (cm³/s)
600
500
400
300
200
100
0
(c)
-200
-100
0
100
TIME FROM RELEASE (ms)

Fig.25


FREQ (kHz)
0
1
2
3
4
5
6
7
8
0
100
200
300
400
500
600
TIME (ms)

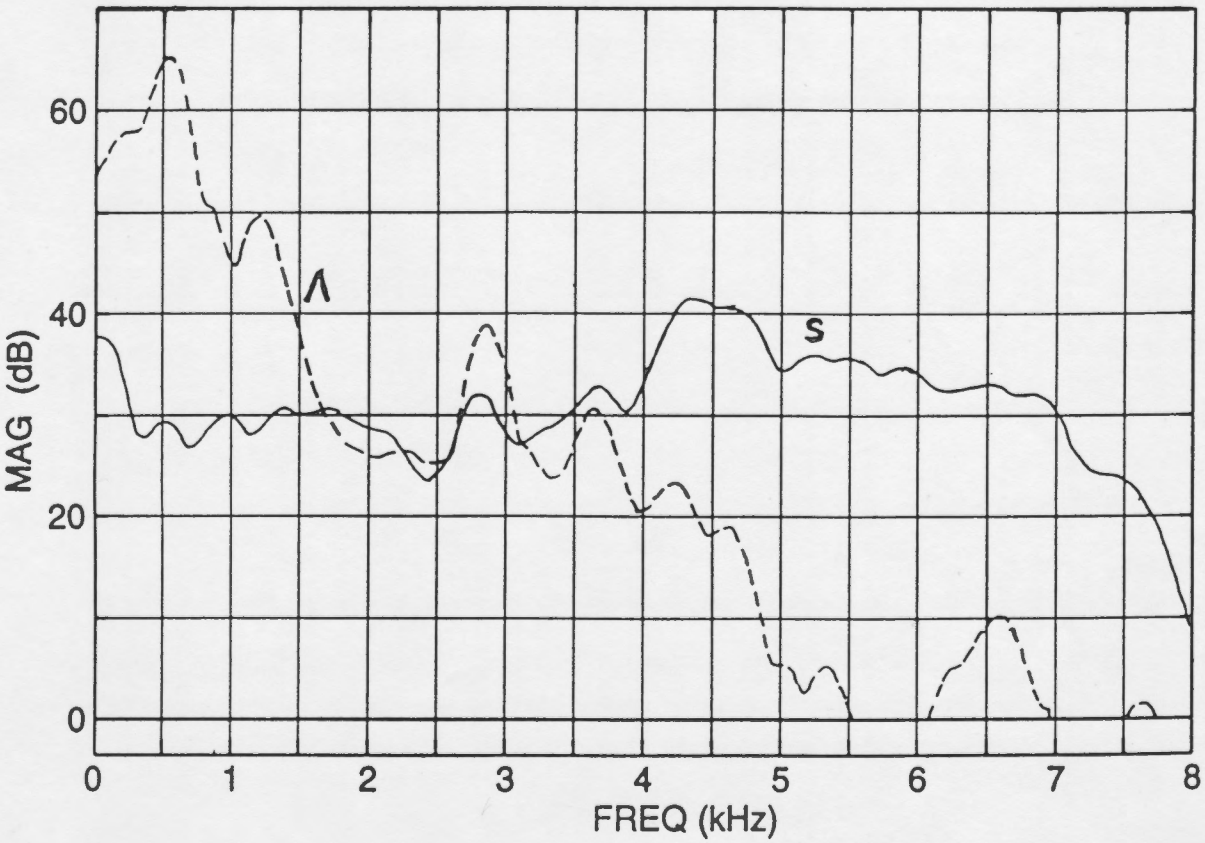

MAG (dB)
60
40
20
0
Λ
S
0
1
2
3
4
5
6
7
8
FREQ (kHz)

Fig. 26

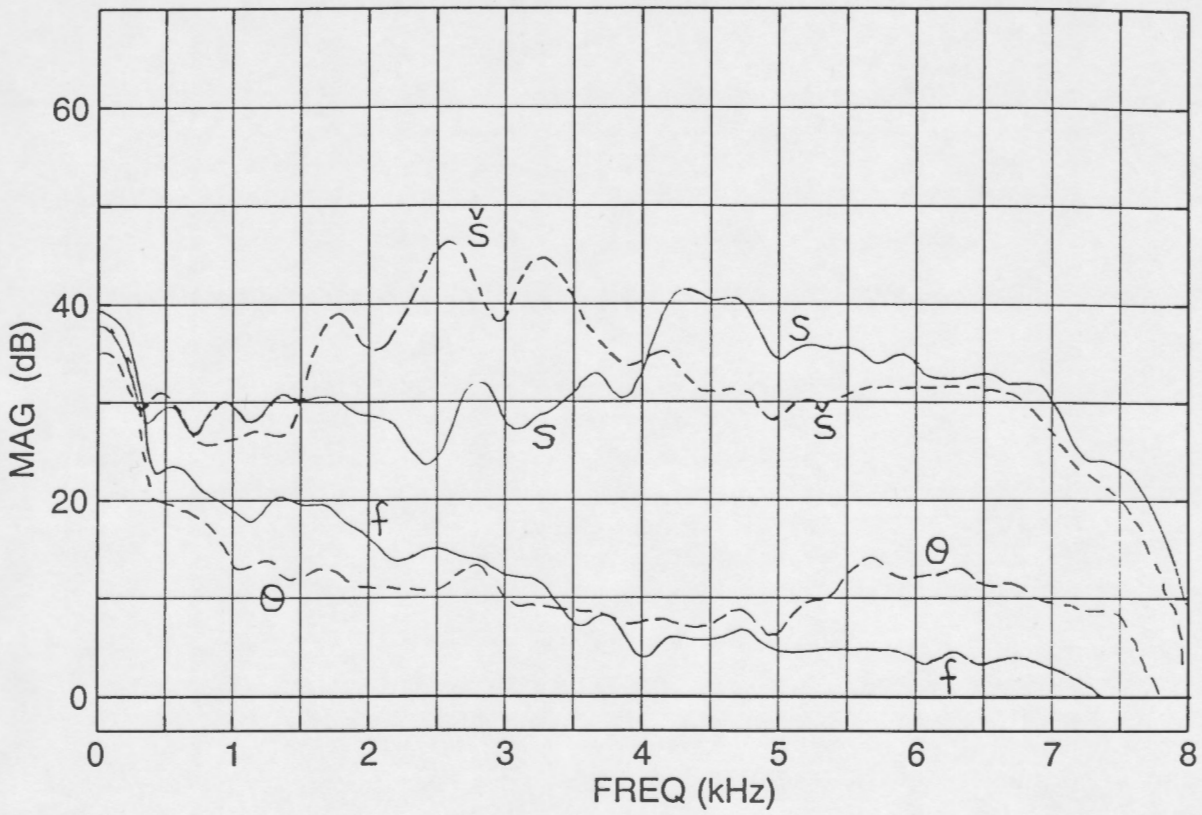


Fig.27

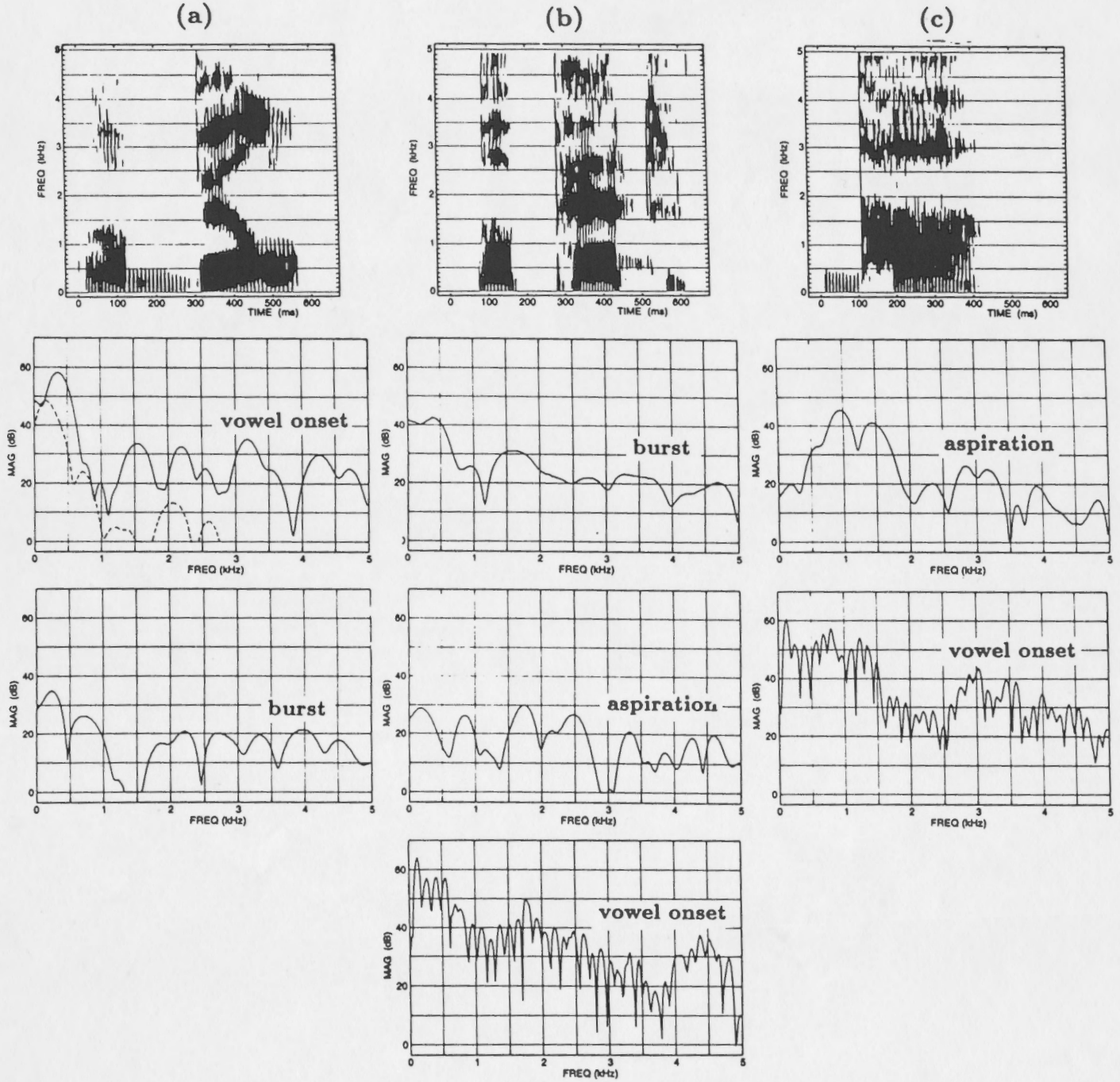

(a)
(b)
(c)
FREQ (kHz)
TIME (ms)
vowel onset
burst
aspiration
MAG (dB)
FREQ (kHz)

Fig. 28