INTRODUCTION
to
COMMUNICATION SCIENCES
and
DISORDERS

FRED D. MINIFIE

CHAPTER **10**

Scientific Substrates of Speech Production

KENNETH N. STEVENS, PH.D.

After reading this chapter, you should be able to:

- Describe an utterance as a sequence of words, and describe the words as sequences of phonetic units.
- Interpret the production of a speech sound in terms of the generation of sound sources and the shaping or filtering of these sources.
- Recognize that a sound can be described as a waveform that changes with time or as a spectrum that represents the sound as a sum of sinusoidal waveforms.
- Describe how sound is generated by vibration of the vocal folds.
- Describe how sound is generated by turbulence in the flow of air at an obstacle or constriction in the airway.
- Interpret vowel sounds in terms of vocal tract shapes and acoustic characteristics.
- Classify consonant sounds in terms of how they are produced in the vocal tract and in terms of their acoustic properties.

A theory of speech production is reviewed, with emphasis on the mechanisms of sound generation in the vocal tract. The sound that emerges from the lips is the result of a two-step process: the generation of sound sources that are the result of modulation of airflow through a constricted part of the airway between the larynx and the lips and the filtering of these sources by the acoustic cavities of the vocal tract. The acoustic properties of a number of different classes of sounds are described, including influences on the sounds when they occur in sequences of words and in casual speech.

As an initial step in the generation of an utterance, a speaker must generate a thought, select a set of words to be produced, and organize these words into a proper sequence. Once this sequence of words has been laid out or planned, instructions are given to the **articulators** (movable structures within the vocal tract or airway extending from the larynx to the mouth opening) and to the respiratory system to perform an appropriate sequence of movements. In his or her memory, the speaker has access to a list of words or a lexicon. Each word in this lexicon contains a set of instructions that specify how the various articulators are to be maneuvered to produce that word. These instructions for each word can be organized to produce a sequence of sounds or **segments**. This pattern of organization provides specific instructions for moving the articulators that are crucial to the production of each of the component sounds. The movements of the articulatory structures specified by the lexicon result in the generation of sound that is radiated from the mouth, nose, and neck surfaces of the speaker.

The primary articulators that are active in producing different speech sounds are shown in Figure 10-1. This figure is a section through the midline of the head, and it shows the trachea or windpipe, the **larynx** located above the trachea, and the structures that shape the airway above the larynx. During speech production, the primary function of the respiratory system below the trachea is to provide a supply of air that is converted to sound in the vicinity of the larynx or in the airway downstream (toward the mouth opening) from the larynx. The articulators that can be manipulated to produce different sound patterns during speech include the vocal folds, which are located within the larynx, structures in the pharyngeal region, the soft palate, the tongue body, the tongue blade, and the lips. The function of each of these structures in producing different sounds will be discussed in the following sections.

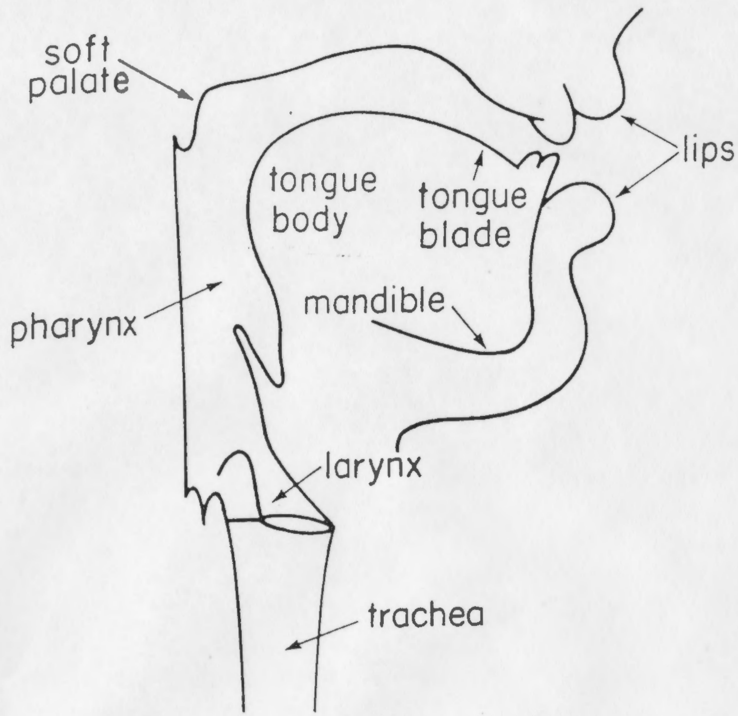

Figure 10-1. Sketch of the shape of the human vocal tract in the midline. The various parts of the system that are active in speech production are identified.

✦ SOUND SOURCES AND SOUND FILTERS IN SPEECH PRODUCTION

The generation of most sounds, including the sounds of speech, can be described in terms of two processes: (1) generation of a **source** of sound energy and (2) a **filtering** or shaping of this source. This distinction between a sound source and a filter can be illustrated by examining the sound generated by a drum beat. The source in this case is the sound produced by the initial tap of the drumstick on the drumhead. This source sets the stretched drumhead in motion, and the character of the sound that emerges depends on the shape and tension of this surface. The drumhead acts to filter the initial brief pulse of sound energy imparted by the drum beat.

As another example, a brief sound source is generated when a guitar string is plucked. In this case, the filtering is dependent on the tension and mass of the string, where along its length the string is plucked, and the configuration of the body of the guitar. For some musical instruments, such as the violin, the source of sound is continuous rather than impulsive. The bow is passed over the string, and the sound from this continuous source is filtered by the string and the violin body. The filtering of the sound source in these examples is the result of mechanical properties of a structure (such as the drumhead or the guitar string), but it can also be an acoustic filtering by tubes or resonators, as in woodwind or brass instruments.

A familiar example of a sound produced by an acoustic source and filter is the sound that is generated by blowing over the top of a pop bottle. The flow of air over the neck of the bottle produces a sound source, and this source is filtered by the resonance of the air column in the bottle.

In the production of speech, sound sources are generated by narrowing the airway, either at the larynx or at some location downstream from the larynx, and causing air to flow rapidly through this narrow constriction. When there is little or no narrowing of the airway, as in quiet breathing, the passage of air causes little or no audible sound to be generated. If the narrowing is produced by placing the vocal folds close together, the folds may be set into vibration, and the modulated flow of air through the space between the vocal folds (the glottis) forms the acoustic source. A source also can be generated by creating turbulence in the flow of air in the vicinity of a constriction (e.g., during the production of an /s/ sound). Such sound sources are much like the sound that is produced at the outlet of an air conditioner. This sound source for an /s/ sound has a noisy character that is quite different from the sound that is generated by vocal fold vibration.

The filtering of these sound sources is achieved by shaping the airway between the glottis and the lips. The adjustment of the source-generating constrictions as well as the shaping of the airways for acoustic filtering is accomplished by manipulating the positions and shapes of the various articulators, as discussed above.

✦ CHARACTERIZATION OF WAVEFORMS, SOURCES, AND FILTERS

One way of describing a sound at a point in space is in terms of the time variation of the pressure fluctuations in the air. This variation in pressure can be depicted as a graph of sound pressure versus time. This graph is called a **waveform**. An example of such a graph for a sound in which the pressure variation with time is sinusoidal is shown in Figure 10-2A. The pattern of sound pressure is repeti-

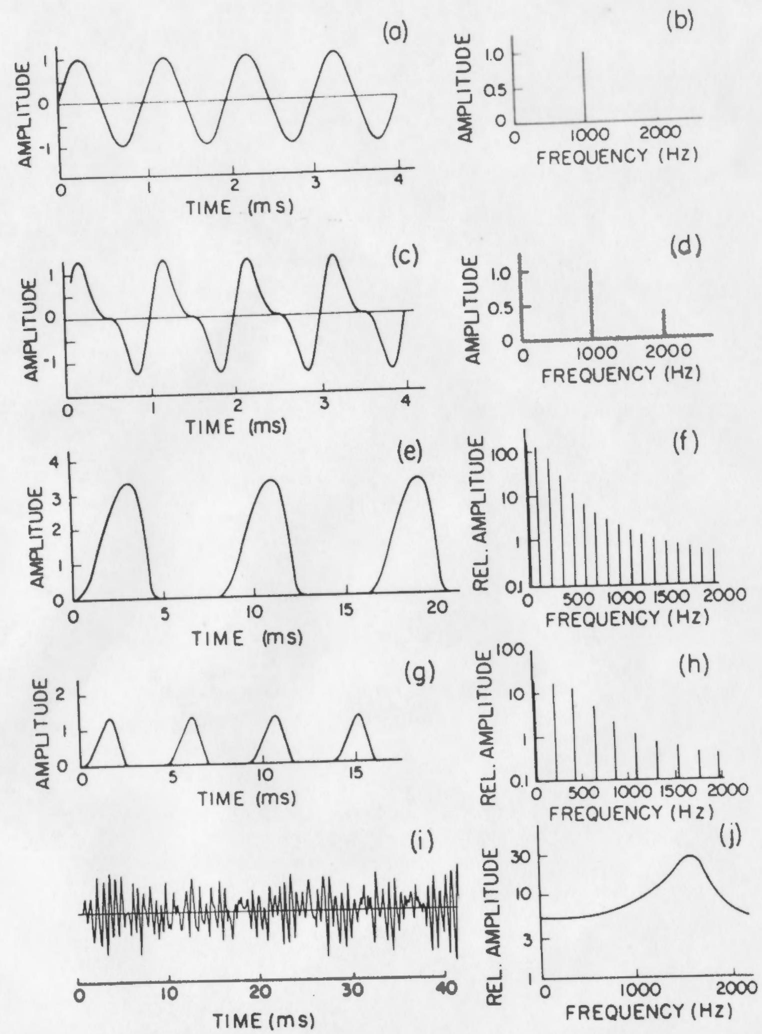

Figure 10-2. Examples of several different waveforms (right panels). The top panels A and B show a sinusoidal waveform, which has just one component in the spectrum, depicted by the vertical line at 1000 Hz. The waveforms in the next three panels C, E and G are periodic, and their spectra D, F and H have several components with equal spacing between the components. The bottom panel I shows a noise waveform for which the spectrum J is continuous.

tive or **periodic** (repeats itself in the same amount of time), and the number of complete cycles per second is the **frequency**. In this example, the frequency is 1000 Hertz (or cycles per second) and the period is 1 millisecond. (This means the waveform repeats itself 1000 times per second.) An alternative way of depicting such a sinusoidal waveform is given in Figure 10-2B. In this representation, the horizontal axis indicates the frequency of the sinusoidal waveform, and the vertical axis indicates the amplitude of the waveform, which in this case is unity. The vertical line at 1000 Hz (abbreviation for Hertz) gives the frequency and amplitude of the sinusoid. This second representation is called the **spectrum** of the waveform. The two representations are equivalent. The time waveform shows a sinusoidal variation with a particular amplitude and frequency. The spectrum is a somewhat more compact representation, which shows the frequency by the placement of the vertical line along the horizontal axis and the amplitude by the vertical length of the line.

Figure 10-2C shows a waveform constructed by adding two sine waves — one at a frequency of 1000 Hz and an amplitude of 1 unit, and the other at a frequency of 2000 Hz and an amplitude of 0.5 unit. This complex waveform (nonsinusoidal) repeats itself every 1 millisecond, and hence is periodic with a frequency of 1000 Hz. The spectrum of this waveform, shown in Figure 10-2D, has two components, as depicted by the two vertical lines.

It turns out that any **complex periodic waveform** can be constructed by adding a set (two or more) of sinusoids whose frequencies are multiples of the frequency of the waveform. The waveform in Figure 10-2C is an example with just two components. Figures 10-2E to H show two more examples. As before, the waveform is given at the left, and the corresponding spectrum is on the right. Each of these waveforms has a number of components, with the amplitudes of the components being smaller as the frequency increases. The sinusoidal component whose frequency is equal to the repetition frequency of the waveform is called the **fundamental** component or the first harmonic, and the remaining components are called **harmonics** (whole number multiples of the fundamental). Note that in these examples the amplitude scales for the spectra are logarithmic. The examples in Figures 10-2E and G have different fundamental frequencies, and hence different spacing between the harmonics. The sounds from many musical instruments are rich in harmonics, and these sounds can also be represented by spectra like those in Figures 10-2F and H.

Some speech waveforms such as the fricative sound shown in Figure 10-2I are not periodic. Such waveforms are called **aperiodic** because they do not repeat themselves. These waveforms also can be represented by spectra, but these spectra have a very large number of closely spaced components. It is convenient to represent the spectra of such sounds by a continuous curve, as in Figure 10-2J. This curve gives the relative amplitudes of the closely spaced frequency components at each frequency.

> The fundamental frequency F (in Hz) is inversely related to the amount of time (the period P in seconds) needed to complete one cycle of the waveform. The equation $F = \frac{1}{P}$ describes this relation. For example, if the waveform repeats itself once every 0.005 sec, then the frequency is
>
> $$F = \frac{1}{0.005} = 200 \text{ Hz, or cycles per second.}$$

There are at least two motivations for representing sounds such as speech in terms of spectra. One motivation is that the ear processes sounds to give an approximate spectral representation of sounds in the neural pathways from the ear to the brain. Thus, when sounds of the type shown in Figure 10-2 impinge on the ear, there is a representation of these sounds inside the ear that is similar to spectral patterns like those shown at the right of the figures.

A second motivation for characterizing waveforms in terms of sums of sinusoidal components comes from the special characteristics of certain "filters," such as the human vocal tract, amplifiers, loudspeakers, organ pipes, and so on. When an input signal is applied to these devices, there is an output signal in response. The special characteristic of these systems is that if the input signal is a sinusoid with a particular amplitude and frequency, then the output is a sinusoid with the same frequency but with a different amplitude. The situation is depicted schematically in Figure 10-3A, B, and C.

Figure 10-3A is an example of a filter characteristic. If the input to the filter is a sinusoid with a particular frequency, the curve in Figure 10-3A gives the amount by which to multiply the amplitude of the sinusoid at that frequency to obtain the amplitude of the output sinusoid. For example, if the input is a sinusoid with a frequency of 1000 Hz, as shown in the spectrum in Figure 10-3B, then the filter characteristic in Figure 10-3A indicates that the amplitude of that sinusoid is multiplied by 3 to obtain the amplitude of the output. Because the amplitude of the input is 1, the output amplitude is 3, as in the spectrum in Figure 10-3C. In general, the amount by which the filter changes the amplitude of the input sinusoid depends on the frequency of the sinusoid. Suppose, for example, that the input signal is periodic but with several frequency components,

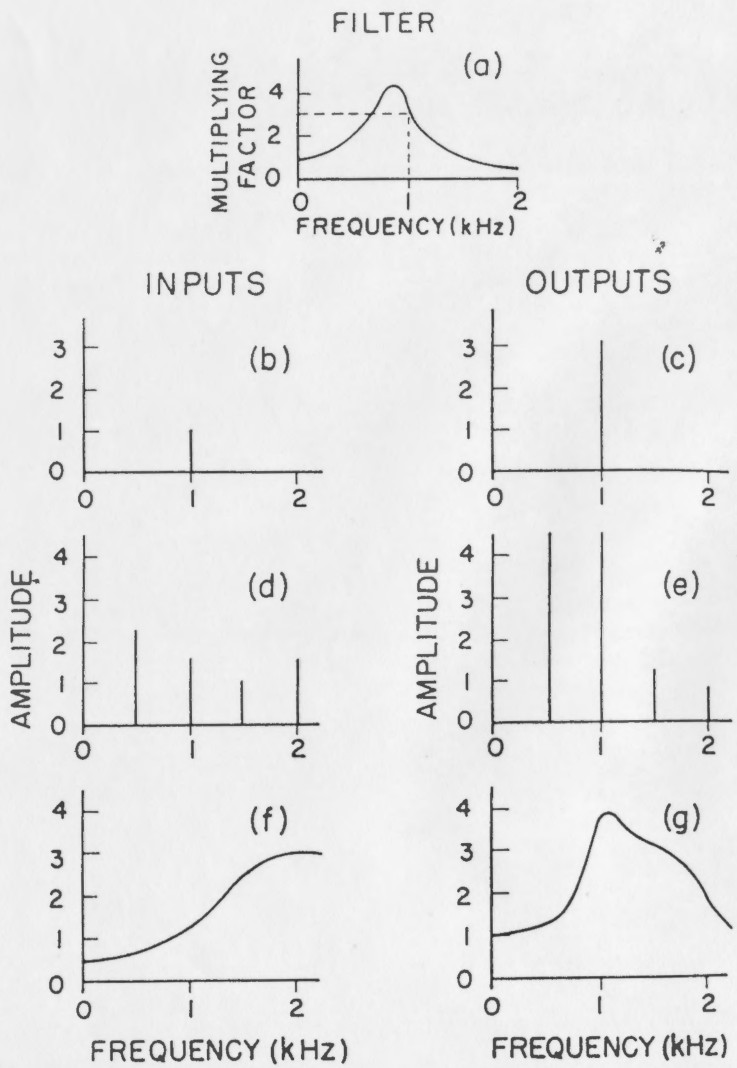

Figure 10-3. Illustration of the filtering of different waveforms. The filter characteristic, shown in A, indicates the factor by which the amplitude of each frequency component of the input is multiplied to obtain the output amplitude of that component. Panels B, D and F show the spectra of three different input signals, and panels C, E and G show the spectra of these signals after they have been passed through the filter.

as in the spectrum shown in Figure 10-3D. Each sinusoid in this spectrum is multiplied by a different factor as it passes through the filter. After passing through the filter, the spectrum of the signal has the form shown in Figure 10-3E. In this example, some of the harmonics in the output signal are larger than those in the input signal, and some are smaller.

If the spectrum of the input signal is continuous, like the spectrum shown in Figure 10-3F, then the output of the filter also has a continuous spectrum, as illustrated in Figure 10-3G. At each frequency, the output spectrum is equal to the input spectrum multiplied by the filter characteristic.

When we consider how different sound sources are generated in the **vocal tract** and how the vocal tract filters these sources, we generally describe the sources in terms of their spectra. We will examine how these sources are filtered by the vocal tract to give the spectrum of the output sound pressure.

✦ VOCAL-FOLD VIBRATION

The source of sound that is used for the generation of vowels and some other speech sounds is produced by setting the **vocal folds** into vibration. The vocal folds are located in the larynx, which is positioned at the upper end of the trachea or windpipe. The position of the vocal folds is shown in Figure 10-4, with their front ends attached to a structure called the thyroid cartilage, and their back ends attached to small cartilages called arytenoid cartilages. The vocal folds are about 1 cm long for an adult female and about 1½ cm for an adult male. Viewed from above, the vocal folds appear as two bands of tissue, about 2-3 millimeters thick, as shown in Figure 10-4.

During normal quiet breathing the vocal folds are set rather far apart, as in Figure 10-4A. The vocal processes of the arytenoid cartilages are rotated toward the side of the airway. This causes a large opening between the vocal folds (glottis). Under this condition very little sound is produced by the gentle inward and outward flow of air. When the vocal folds are positioned close together during exhalation of air from the lungs (as in Figure 10-4B), pressure is built up in the trachea, and the vocal folds are set into vibration.

The series of pictures in Figure 10-5 demonstrates how this vibration occurs. These pictures show how the middle of the vocal folds would appear if they were observed in a coronal cross section (from the front). The series of pictures indicates what would be observed if a rapid series of photographs of the vocal folds were taken during 1 cycle of vibration. At the time of the first frame, the upper edges of the folds remain closed, and the air pressure in the trachea exerts a

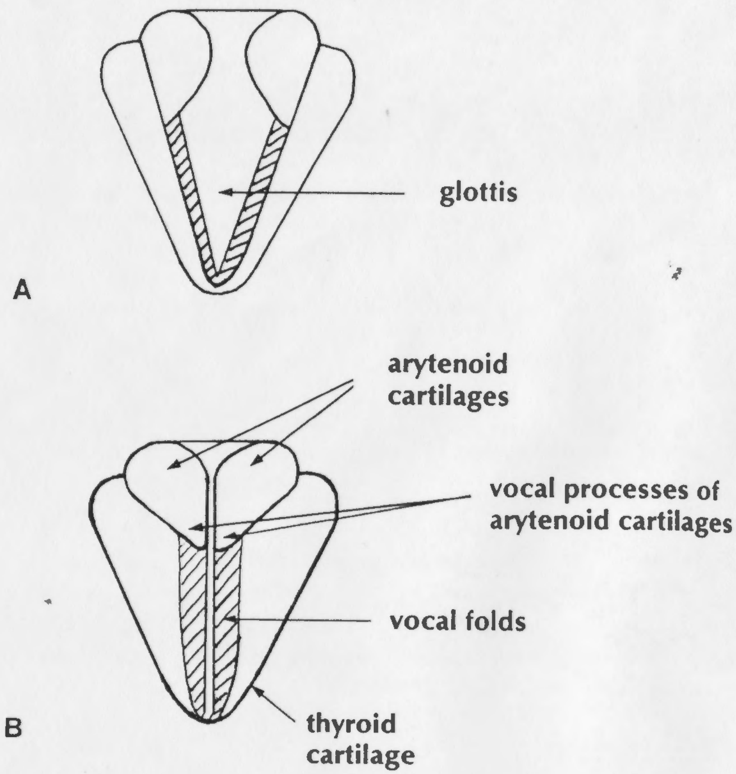

Figure 10-4. Schematized representation of the vocal folds and surrounding structures as seen from above. **A.** Configuration of these structures during quiet breathing. **B.** Configuration when the structures are in a position that leads to vocal fold vibration.

force on the lower parts of the folds, as shown by the arrows. This force pushes the lower surfaces apart (frame 2), until the upper edges open at frame 3. At this time, air begins to flow through the channel between the vocal folds, called the **glottis**. The upper edges of the folds continue to move outwards (frame 4), but because of the flow, the pressure in the space between the folds decreases. Due to the elasticity of the vocal folds, and the rate of airflow through the glottis, the folds spring back together (frame 5), first touching at the bottom and then at the top. The flow of air is cut off and the cycle repeats itself.

As a consequence of this vocal fold vibration, brief pulses of air flow through the glottis (see volume velocity graph in bottom

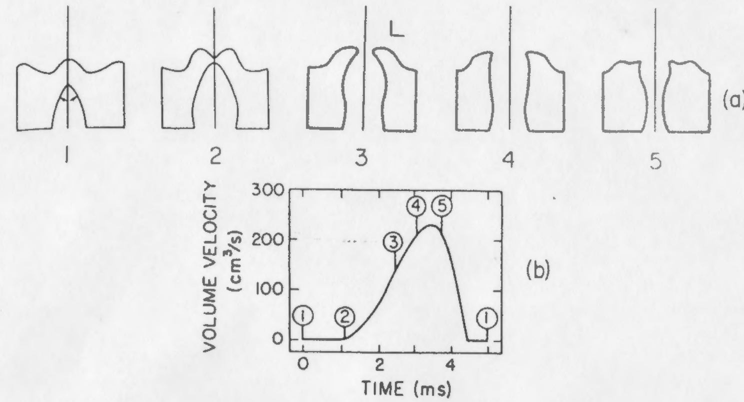

Figure 10-5. A. Schematized coronal sections (i.e., as viewed in a plane perpendicular to the vocal folds) showing the configuration of the vocal folds at various instants of time during a cycle of vibration (adapted from Baer, 1975). Successive times are labeled 1-5, with the **glottis** (the opening between the vocal folds) 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 of air pressure 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.

panel of Figure 10-5). These pulses constitute the acoustic source for vowels and many other speech sounds. The velocity of the flow of air through the glottis looks something like the periodic waveforms that have already been shown in Figures 10-2E and G. The frequency of vibration of the folds, and hence the frequency of the pulses of air, depends on the thickness, mass, and tension of the vocal folds. For adult female speakers, this frequency is usually in the range 150-300 Hz (e.g., as in Figure 10-2G), whereas for adult males, this range is about 90-180 Hz (as in Figure 10-2E). The frequency is somewhat higher for children than for adult females.

The waveform and spectrum of the air pulses through the glottis may vary from one individual to another. Normally the pattern is regular and produces a complex periodic waveform like those in Figure 10-2E and G. Sometimes the vocal folds do not function correctly, and the pattern of pulses emerging from the glottis is irregular and noisy. That is the case with certain voice disorders.

✦ VOWELS

During the production of a phrase or sentence, the various articulators above the larynx are moved around to change the shape of the airway between the glottis and the lips. This tortuously shaped airway is called the **vocal tract**. The sequence of articulatory movements tends to alternate between times when this airway is relatively open and times when a narrowing or constriction is formed by one of the articulators. **Vowel sounds** are produced during the times when the vocal tract is relatively open, whereas consonants are generated during the constricted intervals. We look first at the vowel sounds, for which, as we have seen, the acoustic source is at the glottis.

For vowels, the shaping of the airway by the articulators determines how the sound source at the glottis is filtered. Figure 10–6 shows a sagittal cross section (from the side) of the shape of the airway for four different vowels. This shape is the outline of the surfaces of the vocal tract in the midline, and does not show the dimensions in a lateral direction (away from midline, toward the side). Nevertheless, the midsagittal shapes in Figure 10–6 show how the positions of the tongue body and the lips can be manipulated to produce very different configurations for the airway. The sound from the glottal source is filtered differently for each of these shapes. The filtering is determined by the variation of the cross-sectional area of the airway along its length from the glottis to the lips. For the vowels /i/ (as in *beat*) and /u/ as in *boot*), the tongue body is raised up in the mouth, leaving only a narrow passage between the tongue surface and the palate. The tongue body is displaced farther back in the mouth for /u/ than for /i/, and the lips are more protruded and form a relatively narrow opening at the front of the vocal tract. In the case of the vowel /ɑ/ (as in *father*), the tongue body is low in the mouth, and the mandible or jaw is lower than for /i/ and /u/. The tongue body is displaced back to form a rather narrow opening in the pharyngeal region during the production of /ɑ/. In the case of the vowel /ə/ (as in the first vowel in *about*), the cross-sectional area is approximately the same along the length of the airway.

The filtering produced by the vocal tract for the uniform configuration represented by /ə/ is shown in Figure 10-7. This filter characteristic gives the amount by which each frequency component of the pulses of airflow at the glottis is multiplied to give the amplitude of this component of the velocity of back-and-forth fluctuations of the air at the lips. There are several peaks in the filter characteristic in Figure 10–7, and these peaks are approximately uniformly spaced. At the frequencies of the peaks, the source spectrum is multiplied or amplified by a factor of 5–10, whereas at the valleys

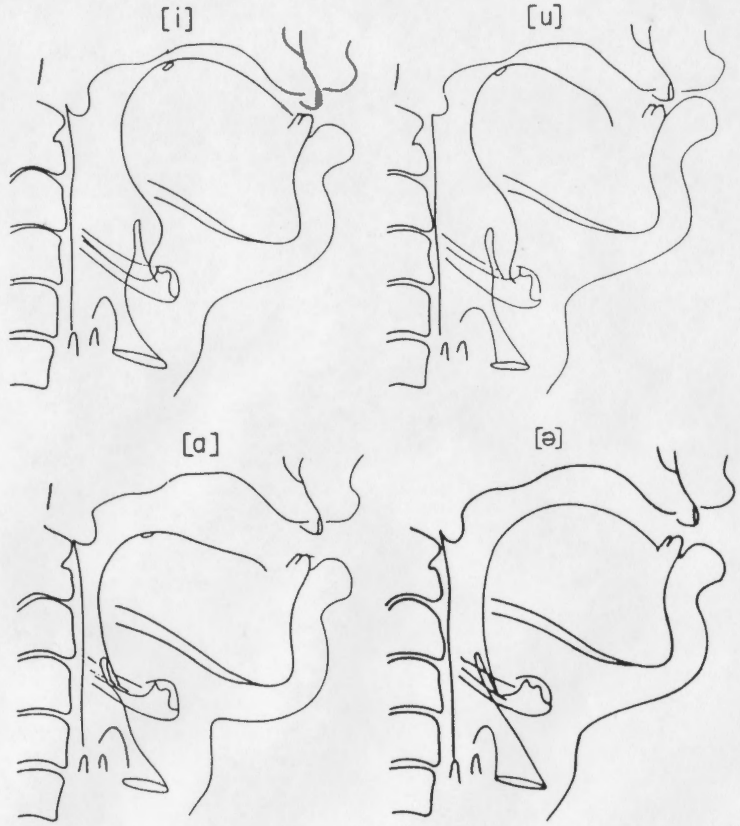

Figure 10–6. Sketches of the outline of the vocal tract at the midline when four different vowels are being produced: /i/ as in *beat*, /u/ as in *boot*, [ɑ] as in *father*, and /ə/ as in the first vowel in *about*. The figures are adapted from illustrations in Perkell (1969).

the multiplication factor is close to unity. The peaks in the filtering are evidence of **resonances** of the vocal tract, called **formants**. The frequencies of the formants depend on the speed of sound and on the overall length of the vocal tract along a path from the glottis to the lips. For this example of a uniform airway with a length l, the formants are located at frequencies of

$$\frac{c}{4l}, \frac{3c}{4l}, \frac{5c}{4l}, \text{ and so on.}$$

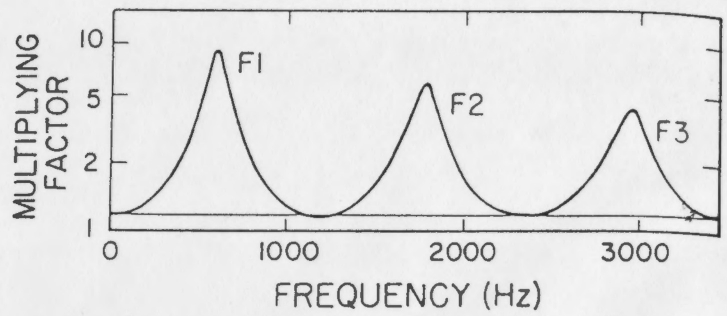

Figure 10-7. Filter characteristic of the vocal tract when it is in a configuration for which the cross-sectional area is uniform along its length. The filter characteristic gives the factor by which the spectrum of the acoustic source at the glottis is multiplied to obtain the acoustic output at the lips. The length of the vocal tract is taken to be 15 cm, corresponding to the average length for an adult female. The peaks corresponding to the first three formants are identified.

where c = speed of sound = 35400 cm/second. Thus, for example, if the length l is 15 cm (corresponding approximately to the average vocal-tract length for an adult female speaker), the first three formant frequencies are 590, 1770, and 2950 Hz. as shown in Figure 10–7. The average vocal-tract length for an adult male speaker is about 17 cm, leading to formant frequencies of 520, 1560, and 2600 Hz for this uniform configuration.

As we have observed, the result of filtering the spectrum of the glottal pulses by the filter characteristic in Figure 10–7 is the spectrum of the velocity fluctuations at the lips. There is some additional modification of these frequency components at the lips to account for radiation of the sound, leading to the spectrum of the sound pressure at a distance from the speaker's mouth. The steps in the production of the sound, from the glottal source to the sound pressure at a specific distance from the lips, are summarized in Figure 10–8. At the initial step, both the waveform $U_s(t)$ and the spectrum $U_s(f)$ of the source are shown. This source is filtered by the vocal tract and by the radiation, and the result is the waveform $p(t)$ and spectrum $P(f)$ of the sound pressure, as displayed at the right of the figure.

When the shape of the vocal tract is not uniform (e.g., for the vowels /i/, /u/, and /ɑ/ in Figure 10–6), the filtering component of this process in Figure 10–8 is modified. The filter still shows evi-

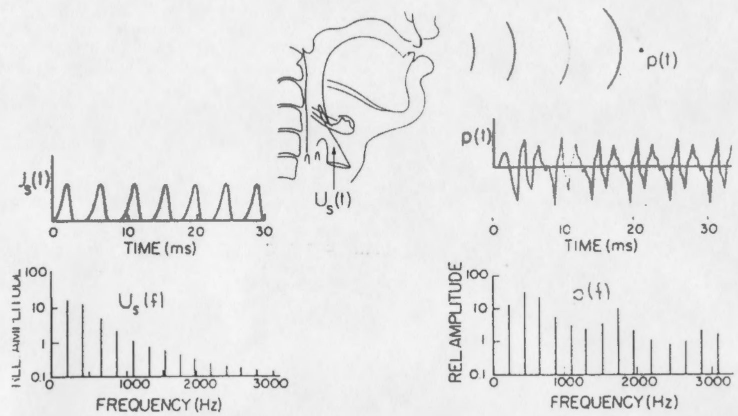

Figure 10-8. Showing how the acoustic source $U_s(f)$ at the glottis is modified by the vocal tract and by radiation from the lips to obtain the sound pressure $p(t)$ at some distance from the lips. At the left of the figure both the source waveform and its spectrum are shown. The waveform and spectrum of the sound pressure at a specified distance outside of the mouth are given at the right.

dence of prominences or formants, but these formants no longer have uniform spacing. The first two formants in particular are displaced to higher or lower frequencies depending on how the airway is shaped by the tongue body and the lips. For example, when the tongue body is raised (as for /i/ and /u/ in Figure 10–6), the frequency of the first formant moves downward from its value for a uniform configuration, and when the tongue body is lowered (as for /ɑ/) it moves upward. Lowering of the tongue body usually is accompanied by lowering of the jaw or mandible, leading to a vocal tract shape that is more open in the front. On the other hand, forward and backward movements of the tongue body in the mouth displace the frequency of the second formant toward a higher and lower frequency, respectively.

The frequencies of the first and second formants for several different vowels are plotted in the chart in Figure 10–9. The vowels represented in this chart are the four vowels whose vocal-tract shapes are displayed in Figure 10–6, together with the vowel /æ/ as in the word *bat*. The chart is arranged so that the vowels produced with a high tongue-body position (and a low first-formant frequency) are at the top, and those produced with a low tongue-body position (high first-formant frequency) are at the bottom. Front vowels are at the right and back vowels at the left.

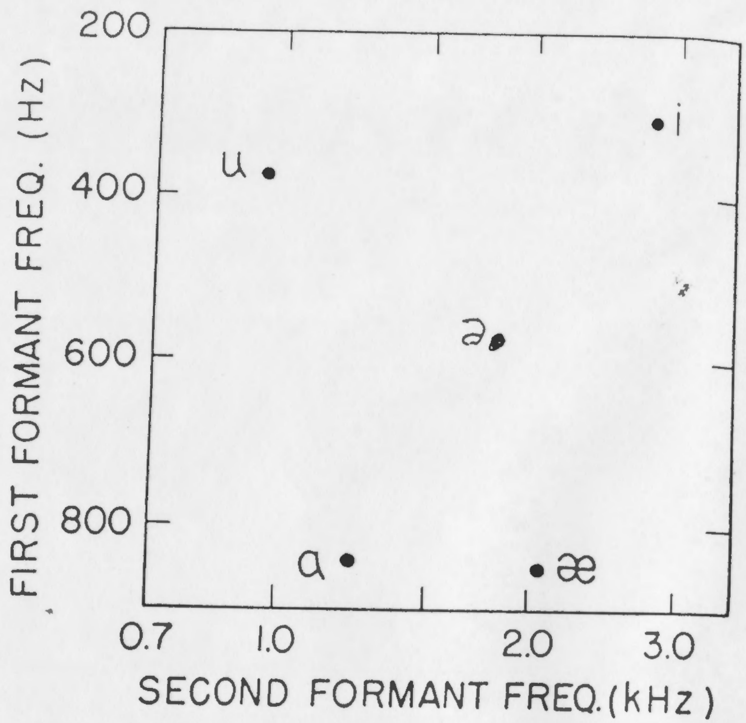

Figure 10-9. The frequency of the first formant is plotted against the frequency of the second formant for five different vowels of American English. These are average values for adult female speakers.

In English, there are several vowels in addition to those shown in Figures 10-6 and 10-9. The first two formant frequencies for these vowels generally lie on or within the contour joining the four outer vowels in Figure 10-9. Several of the vowels in English are produced with a vocal tract shape that changes with time, so that the frequencies of the formants are moving through the duration of the vowel. These vowels are called **diphthongs** (vocalic glides within a single syllable). Examples of diphthongs are the vowels in the words *bide, shout*, and *void*.

Changes in the spectrum of speech sounds over time are often displayed in the form of a spectrogram, as illustrated in Figure 10-10. A **spectrogram** is a plot of frequency versus time, with spectrum amplitude represented as degree of blackness of the pattern. The vowel formants are evident as dark bars on the spectrogram. These bars are relatively horizontal or fixed in frequency for vowels that

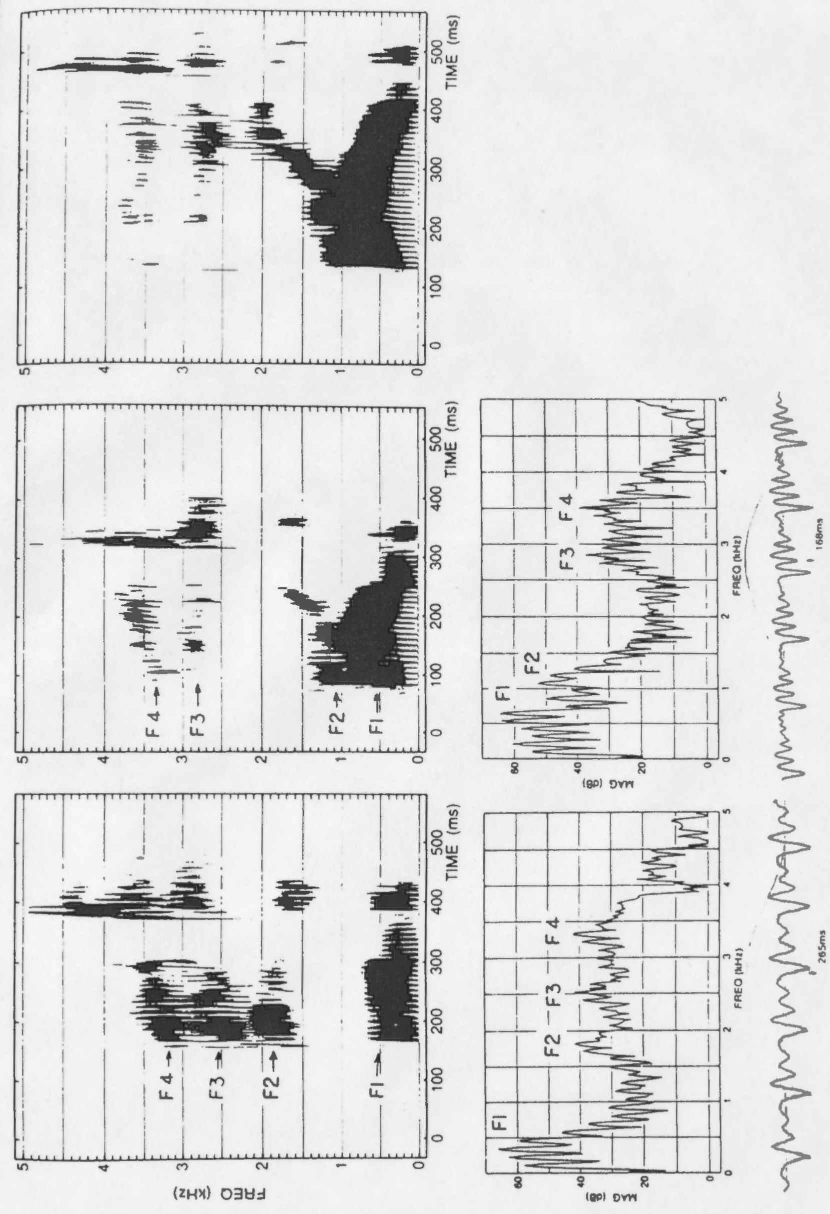

Figure 10-10. The top panels are spectrograms showing the acoustic patterns for words containing three different vowels: *bid* (left), *bud* (middle), and *bide* (right). The vowel in the right panel is a diphthong. The spectra below the first two spectrograms are measured in the middle of the vowels [ɪ] and [ʌ]. A section of the vowel waveform is shown below each spectrum The formant frequencies are identified in the spectra and spectrograms of [ɪ] and [ʌ].

are not diphthongs, but show substantial movement for diphthongs. Vowels of both kinds are displayed in Figure 10–10. For the type of spectrogram shown in Figure 10–10, the regular vertical striations are evidence of the individual glottal pulses that occur when the vocal folds are vibrating. For example, in the spectrogram of the word *bid*, these pulses are spaced about 8 milliseconds apart near the middle of the vowel produced by a male talker, indicating that the frequency of vocal fold vibration is 125 Hz at this point.

Also shown in Figure 10–10 are spectra measured around the middle of the vowels in the words *bid* and *bud*. The individual harmonics are evident in these spectra, and the effects of the formants are observed at frequencies where there are peaks in the spectrum envelope drawn through the tips of the harmonics. The amplitude scale for displaying these spectra (as well as the spectra in later figures) are logarithmic, similar to the amplitude scales used for plotting spectra in Figure 10–8. The vertical axis for these spectra is in decibels (or dB). If a particular spectral component has an amplitude **A**, then its amplitude expressed in decibels is

$$20\log_{10}A \text{ dB}$$

Thus if A is multiplied by a factor of 10, the increase on the vertical scale of Figure 10–8 is 20 dB, whereas for a multiplying factor of 100, the increase is 40 dB.

Figure 10–8 depicts vowel production as the generation of a periodic source at the glottis and the filtering of this source by a vocal tract that has well-defined resonances. These resonances lead to several peaks in the filter characteristic of the vocal tract. This simple picture is sometimes modified, depending on the speaker and on the influence of other sounds that are produced adjacent to the vowel. For example, there may be irregularities in the glottal source, so that the glottal pulses are not completely periodic, or the pulses are accompanied by noise, leading to a voice that sounds rough or breathy. Or, the filtering by the vocal tract may be modified by lowering the soft palate (**velum**). The opening to the nasal cavity that is created by this movement (called the **velopharyngeal opening**) introduces extra prominences into the filter characteristic. (Recall those shown in Figures 10–7 and 10–8.) In English, a velopharyngeal opening is created in a vowel when that vowel is followed by a nasal consonant such as /m/ or /n/. The acoustic effect of this velopharyngeal opening can be seen if we compare the spectrograms and spectra in Figures 10–10 and 10–11. The vowel in the word *bin* in Figure 10–11 shows evidence of the opening to the nasal cavity, whereas there is no such evidence in *bid* in Figure 10–10. For the vowel in *bin*, there is an extra peak in the spectrum at around 1000 Hz.

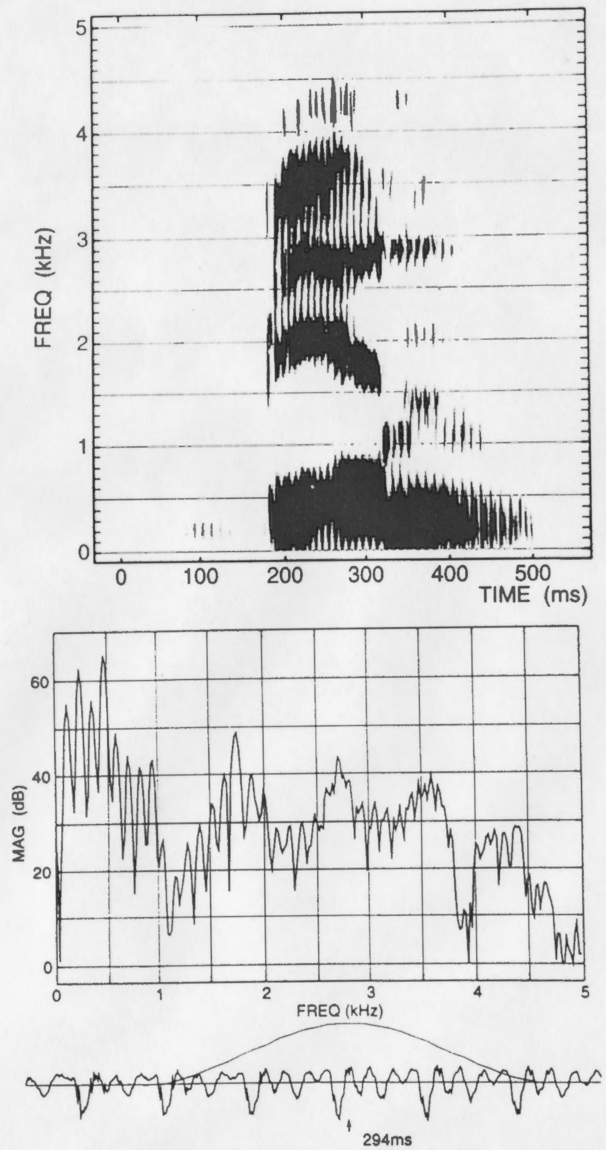

Figure 10–11. The spectrogram at the top is an utterance of the word *bin*. The spectrum in the lower panel is sampled in the vowel of this word. Comparison of this spectrum with that of the vowel [ɪ] in the word *bid* in Figure 10–10 shows some differences, particularly in the frequency range around 1 kHz. The extra peak and valley in the spectrum of the vowel in *bin* is evidence that a velopharyngeal opening has been created in anticipation of the nasal consonant at the end of the word.

✦ CONSONANTS

When vowels are generated, there is only a minor narrowing or constriction in the airway above the glottis, and air flows freely through the vocal tract. **Consonants**, on the other hand, are produced by creating a major constriction in some region of the vocal tract. This constriction is formed by moving one of the articulators (usually the lips, the tongue blade, or the tongue body) so that there is complete closure of the vocal tract, or a very narrow opening in the airway between that articulator and the outer surface of the vocal tract (usually the upper lip or incisors, the **hard palate** [bony roof of the front of the mouth], or the **soft palate** [muscular portion of the roof of the mouth — in the back]).

In English there are about 25 different consonants. These consonants can be classified in terms of: (1) which articulator is used to form the constriction; (2) whether the narrowing produced by this articulator forms a complete closure or only a partial closure of the vocal tract; (3) whether significant pressure is built up in the mouth when the constriction is formed; and (4) how the vocal folds are adjusted when the consonant is being produced. For example, these different properties for the consonant /b/ are: (1) the lips form the articulator, (2) the closure is complete, (3) there is pressure build-up, and (4) the glottis is adjusted so that vocal fold vibration is facilitated. By modifying one of these properties at a time, the consonant /b/ can be changed to a different consonant. Changing the articulator from the lips to the tongue blade (property 1) changes /b/ to /d/; forming a partial rather than a complete closure (property 2) gives /v/; creating a velopharyngeal opening to prevent build-up of pressure (property 3) changes /b/ to /m/; and modifying the glottal adjustment to inhibit vocal fold vibration produces /p/ rather than /b/.

We turn now to a discussion of each of these classes of consonants.

Sonorant Consonants

Consonants that are produced with a narrowing in the vocal tract but with no pressure build-up behind this narrowing are called **sonorant consonants**. The source of sound for these consonants is at the glottis. Because there is no impediment to the flow of air above the glottis, the vocal folds vibrate in much the same way as they do for vowels. Examples of spectrograms of several sonorant consonants as they occur in intervocalic position are shown in Figure 10-12. For all of these consonants, the continuing vertical striations provide evidence of vocal-fold vibration throughout the interval.

Figure 10–12. Spectrograms of several sonorant consonants in the phonetic environment [a–a].

Sonorant consonants differ from vowels in at least two ways. One is that the filtering that occurs in the vocal tract while the tract is constricted has characteristics different from those for vowels. Additional resonances may appear in the spectrum, or some resonances may be suppressed (as in the nasal consonants in Figure 10–12) or two resonances may come close together, merging into a single spectral peak (as in /w/ and /j/ in Figure 10–12). (The symbol /j/ represents the initial consonant in the word *yacht*.) A second attribute of sonorant consonants is that they are always produced with movement of the articulators toward or away from a constricted configuration; hence they always exhibit time-varying spectral characteristics. The movement is usually toward or away from a vowel. Sometimes the time variation is rapid or abrupt (as for the nasal consonants /m, n, ŋ/ or the consonant /l/ in Figure 10–12) and sometimes it is slow (as in /w, j, r/). The abrupt changes in the spectrum for the nasal consonants occur because a complete closure is formed by one of the articulators: the lips for /m/, the tongue blade for /n/, and the tongue body for /ŋ/. (The symbol /ŋ/ is the final consonant in the word *sing*.)

Immediately preceding the closure or immediately following the release for one of these nasal consonants, there is a time interval of about 50 milliseconds in which the vocal tract is moving from the preceding vowel toward the consonant configuration or from the consonant configuration toward the following vowel. As the shape of the vocal tract changes during these time intervals, the formant frequencies move in ways that indicate which of the three articulators (lips, tongue blade, or tongue body) is forming the constriction. These differences in the **formant transitions**, particularly for the second and third formants, can be seen in the spectrograms for these three consonants in Figure 10–12. Experiments have shown that listeners make use of these formant transitions in determining which of the three consonants was produced by a speaker.

Obstruent Consonants

Obstruent consonants are produced when the consonantal constriction causes pressure to build up in the vocal tract behind the constriction. There are two acoustic consequences of this build-up of pressure in the vocal tract. As the pressure in the vocal tract increases, the difference between the pressure in the trachea below the glottis and the pressure in the vocal tract above the glottis decreases. This decreased pressure across the glottis results in a reduced amplitude of the glottal air pulses, and with a sufficient increase in the pressure in the vocal tract, the vocal folds may stop vibrating altogether. A second consequence of the increased pressure in the vocal tract behind a constriction is that a rapid flow of air can occur through the constriction. This airflow can become turbulent, and result in the generation of noise in the vicinity of the constriction. Familiar examples are the hissing sound when the consonant /s/ is produced or the short burst of noise that is generated when the lips are opened at the beginning of a word like *pat*. The **turbulence noise source** has a continuous spectrum of the type illustrated in Figure 10–2J. The source is filtered as the sound passes through the vocal tract that is downstream from the constriction. There is a greater variety of obstruent consonants than sonorant consonants in English and in most languages. As noted earlier, these consonants can be produced with different articulators (different places of constriction), with different voicing characteristics, and with complete or partial closure at the constriction. Spectrograms of a number of obstruent consonants as they occur in intervocalic position are shown in Figures 10–13, 10–15, and 10–16.

The consonants in Figure 10–13 are **stop consonants**, for which the articulator forms a complete closure. These consonants are produced by forming the closure with the lips (/b/), the tongue blade (/d/), and the tongue body (/g/). The shape of the vocal tract in the midline when the closure is formed for each of these consonants is shown in Figure 10–14. In the case of the voiced stop consonants, the vocal folds continue to vibrate through all or most of the time interval when there is a closure. This vocal fold vibration can be seen as vertical striations near the bottom of the spectrograms in Figure 10–13 in the interval between the two vowels. Immediately following the time when these stop consonants are released (closure ceases as articulators move toward positions required for the following vowel sound), there is a rapid flow of air through the rapidly increasing opening. This flow of air creates a brief burst of noise, which acts as a source of sound. This source is filtered by the airway downstream from the constriction. The burst of sound that is radiated from the lips, therefore, has a different spectrum depending on the place in the mouth where the constriction is formed.

Spectra of the bursts for the three consonants /b, d, g/ are shown at the right side of Figure 10–13. The spectrum of the burst for /g/ has a prominence at a frequency of about 1750 Hz, reflecting the resonance of the portion of the vocal tract in front of the constriction formed by the raised tongue body. For the consonant /d/, there is a spectral prominence at a much higher frequency (about 4000 Hz), because the cavity in front of the constriction is relatively short (1.5–2 cm), as shown in Figure 10–14. There is no significant prominence in the spectrum of the burst for the labial consonant /b/, because there is no cavity in front of the constriction to filter the source.

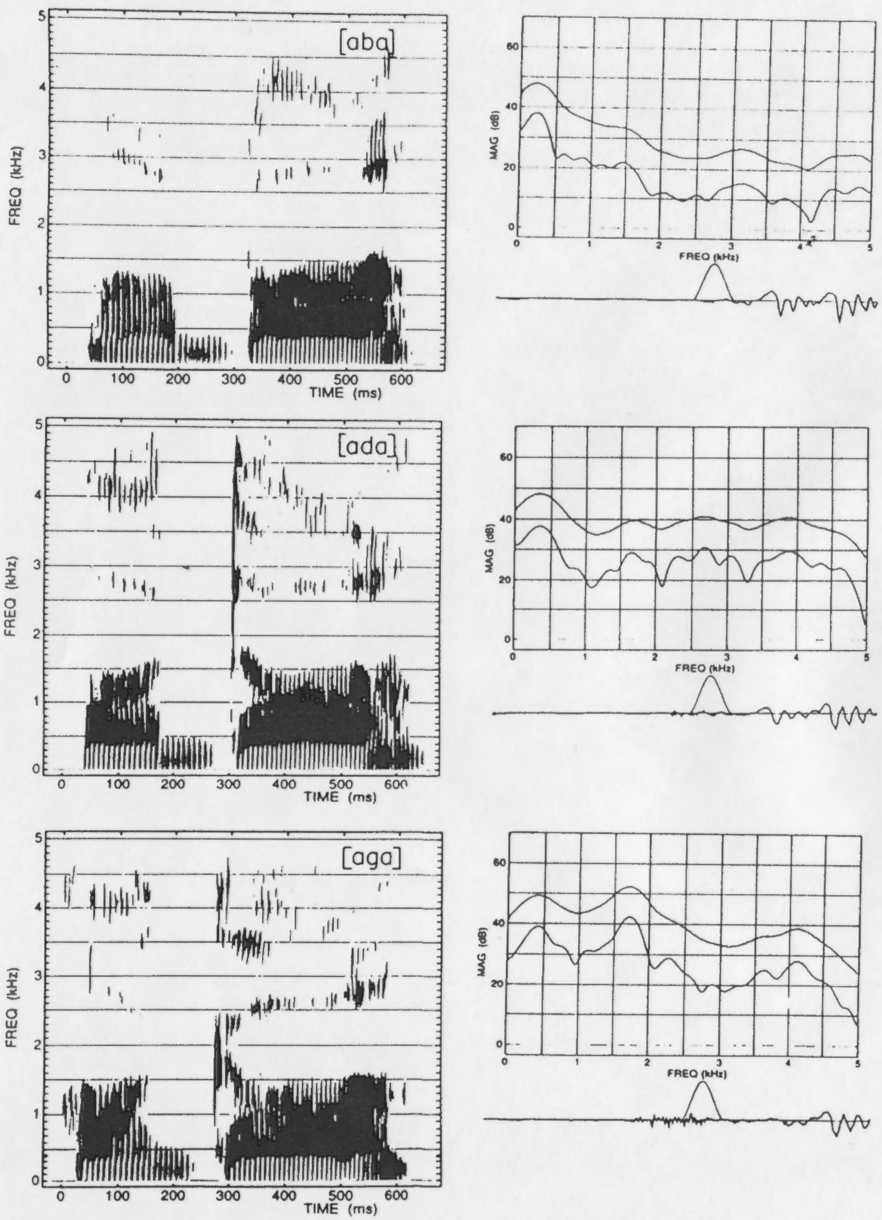

Figure 10–13. The left panels show spectrograms of three voiced stop consonants in the phonetic environment [ɑ-ɑ]. The spectra in the right panels (*lower lines*) are sampled on the release bursts of the consonants, with the waveforms displayed below the spectra. The upper line in each panel is a smoothed version of the spectra. The different spectrum shapes reflect the different positions of the articulators at the time they are released to generate the bursts.

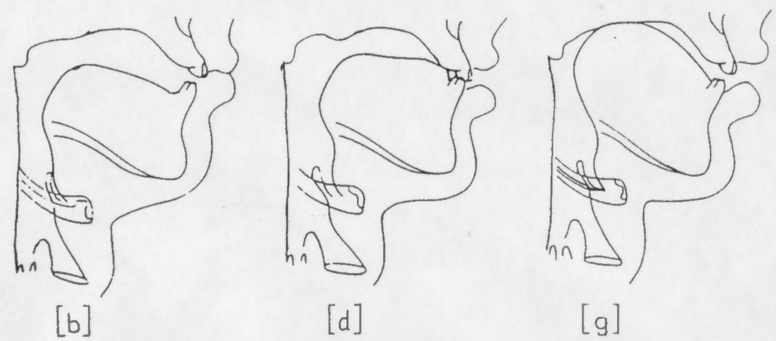

Figure 10–14. Sketches of the outline of the vocal tract in the midline when the closure is formed for three different stop consonants: /b/ (*left*), /d/ (*middle*), and /g/ (*right*). (From Perkell, 1969.)

As with the nasal consonants discussed above, the pattern of change of the formant frequencies (particularly the transitions of the second and third formants) immediately preceding or following the closure interval for the stop consonant provides information to a listener as to whether the consonant was produced with the lips, the tongue blade, or the tongue body. The different movements of the second and third formants just before the consonant closure and just after consonant release are evident in the spectrograms for the three stop consonants in Figure 10–13. This information supplements the information available in the properties of the burst.

For the stop consonants /b, d, g/, the noise burst at the consonant release is relatively short (5–20 ms), and vocal fold vibration occurs immediately after the burst. The vocal folds are already in a configuration such that they begin to vibrate as soon as pressure in the mouth is released and air begins to flow through the glottis. In the case of the voiceless stop consonants /p, t, k/, the vocal folds are spread apart at the time the consonant is released. A spectrogram of the consonant /t/ as it occurs between vowels is shown in Figure 10–15. When air begins to flow through the glottis after the consonant is released, the vocal folds are sufficiently far apart that vibration is not initiated immediately. Around the time of release, the speaker begins to bring the vocal folds together so that they are in a configuration that permits vibration to occur. This movement of the vocal folds takes about 50 milliseconds, and as a consequence, there is a delay of about 50 milliseconds from the time of release to the onset of vocal fold vibration. (This is called the **voice onset time**, or VOT.) During this time interval, turbulence noise is generated, first in the vicinity of the constriction formed by the lips, tongue blade, or tongue

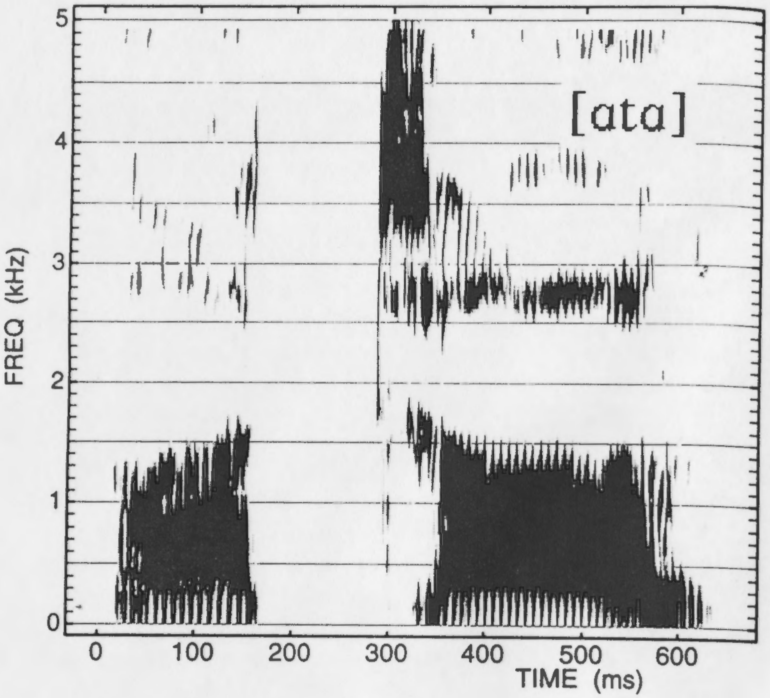

Figure 10–15. Spectrogram of the voiceless stop consonant [t] in the phonetic environment [ata].

body (burst), and then in the vicinity of the glottis (**aspiration**). As was true of the voiced stop consonants /b, d, g/, the spectra of the noise bursts and aspirations reflect the places of articulation for /p, t, k/, namely at the lips, tongue blade, or tongue body. This delay time with noise can be seen in the spectrograms in Figure 10–15. The beginning of the noise (at a time of 270 ms on the spectrogram) occurs well in advance of the onset of vocal fold vibration at about 320 milliseconds.

The spectrograms in Figure 10–16 show examples of the **fricative consonants** /s/ and /z/ in intervocalic position. For these consonants, a narrow constriction is formed by the tongue blade, which is grooved at the midline and placed against the front part of the hard palate. The rapid airflow through the narrow air channel at the constriction creates a turbulence noise source which is shaped or filtered by the short cavity in front of the constriction. This filtering enhances the high frequencies for these consonants, as the spectrograms show. During the time interval in which the noise is generated, there is no vocal fold vibration for /s/. Glottal vibration contin-

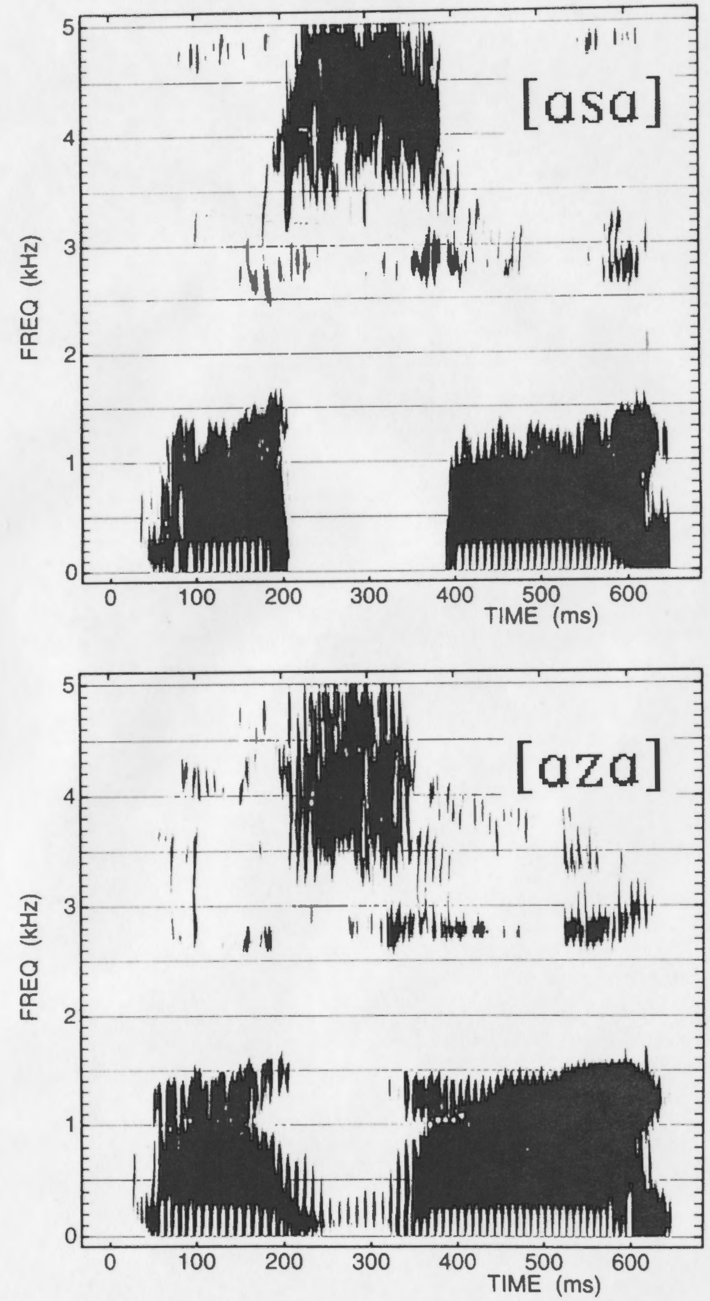

Figure 10–16. Spectrograms of the voiceless and voiced fricatives [s] and [z] in the phonetic environment [a-a].

ues through at least part of this interval for the voiced fricative /z/. Other fricative consonants (/ʃ/ as in *shoe*, /ʒ/ as in *measure*, /f/ as in *face*, and /v/ as in *vase*) are produced by forming the narrow constriction at other places along the front part of the vocal tract. As can be observed by saying these words aloud, each fricative pair has cognate voiced and voiceless consonants at each place of constriction.

In English there are two obstruent consonants that have the characteristics of both stop consonants and fricative consonants. These are the **affricate consonants** /tʃ/ and /dʒ/, as in the words *church* and *judge*, respectively. These consonants are produced by forming a complete closure of the airway by placing the tongue blade against the hard palate, similar to the closure achieved on the stop consonants /t/ and /d/. However, during the production of affricates, when the tongue tip is released, the part of the blade behind the tip continues to form a narrow constriction, and turbulence noise is generated for a time immediately following the release. This latter phase of the affricate release is similar to a fricative consonant.

All of the consonant sounds described above are produced by manipulating the lips, the tongue blade, or the tongue body to form a constriction in the vocal tract in some region above the glottis. The consonant /h/ as in *hat* is an exception, because no narrow constriction is formed in the vocal tract above the glottis. This consonant is produced by spreading the vocal folds apart so that they do not vibrate. When air flows through this relatively spread glottis, turbulence noise is generated in the vicinity of the glottis, and this noise constitutes the acoustic excitation for the vocal tract. A spectrogram of the utterance /ɑhɑ/, shown in Figure 10–17, illustrates the kind of noise that is generated for this /h/ sound. In contrast to fricative consonants or noise bursts for stop consonants, /h/ shows spectral peaks corresponding to several of the vocal tract resonances, and therefore has some of the attributes of a vowel. Because the tongue is not required for the production of /h/, talkers tend to anticipate the position the tongue needs to assume for production of the following vowel. This causes the vocal tract to filter the noise source in a manner similar to that of the following vowel.

Summary of Classification of Consonants

Linguists have developed a framework and terminology for classifying the sounds that are used to distinguish between words in a language. The sounds are classified in terms of a number of properties or features. A partial list of the features that are used to classify consonants is given in the first column of Table 10–1. The remaining columns indicate, for each of a number of different consonants, wheth-

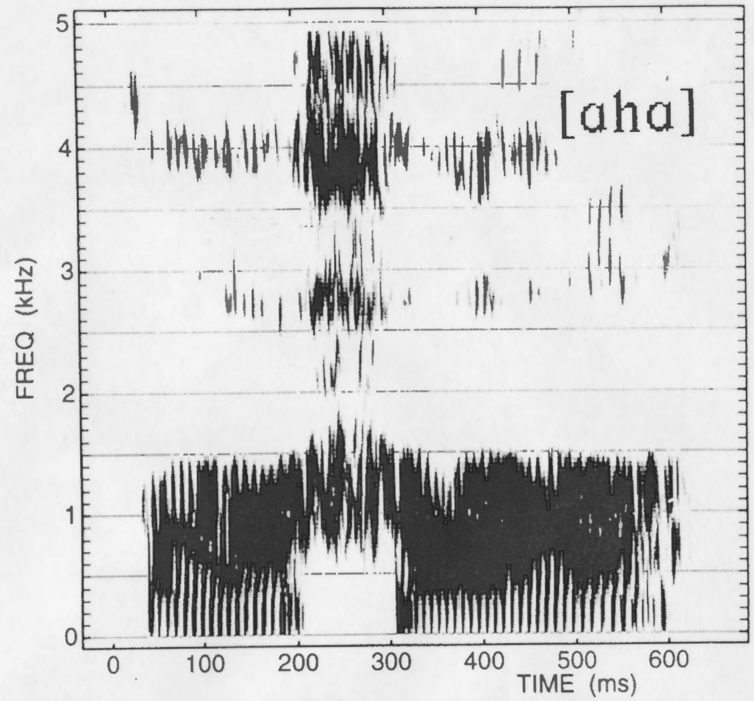

Figure 10–17. Spectrogram of the consonant [h] in the phonetic environment [ɑhɑ].

Table 10–1. A partial listing of feature values for some of the consonants in English.

Feature	p	t	k	b	d	g	f	s	v	z	m	n	ŋ
Lips	+	−	−	+	−	−	+	−	+	−	+	−	−
Tongue blade	−	+	−	−	+	−	−	+	−	+	−	+	−
Tongue body	−	−	+	−	−	+	−	−	−	−	−	−	+
Voice	−	−	−	+	+	+	−	−	+	+	+	+	+
Continuant	−	−	−	−	−	−	+	+	+	+	−	−	−
Sonorant	−	−	−	−	−	−	−	−	−	−	+	+	+

er the various features are present or absent in the sound. This list of six features must be expanded to about 15–20 features to permit classification of all the vowels and additional consonants that can occur in language. See Chapter 3 by Stoel-Gammon and Stone for a more complete list.

The first three features in Table 10–1 specify which of the three articulators is active in forming the constriction in the vocal tract. The feature [**voice**] indicates whether or not the laryngeal structures are adjusted to permit continued vocal fold vibration during the consonant. [**continuant**] indicates whether the articulator forms a complete closure (i.e., [−continuant]) or just a partial closure in the airway, and [sonorant] specifies whether or not pressure builds up in the mouth behind the constriction when the consonant is produced. Pressure build-up occurs for a [−sonorant] consonant. Table 10–1 shows, for example, that the consonant /b/and /p/ share all features except [voice]. On the other hand, for the consonants /n/ and /s/ the features specifying the articulator are the same but the remaining features are different.

Classification of sounds in terms of features also provides an efficient way of describing the constraints on sound sequences that occur in different languages. For example, in English the plural for a noun is formed by appending /s/ to the ends of words like *back* and *laugh*, and by appending /z/ for words like *bed* and *cave*. That is, the [−voice] consonant /s/ is appended if the final consonant of the word is [−voice], and the [+voice] consonant /z/ is appended if the final consonant is [+voice]. Another constraint in English applies when a word ends in an obstruent consonant that is preceded by a nasal consonant, as in *send, lymph*, and *think*. The constraint here is that the nasal segment takes on the same feature [lips], [tongue blade], or [tongue body] that characterizes the final obstruent consonant. A word such as *semt* is not possible in English because it violates this constraint.

✦ SPEECH SOUNDS IN SENTENCES

When phrases and sentences are produced, some of the properties of the sounds that form the words in the sentences may undergo modification, relative to the properties that have been illustrated for the simpler utterances in Figures 10–10 to 10–16. The kinds of modifications depend the speaking style and the context in which the sounds occur. Examples of these influences of context can be seen in the spectrogram of the sentence in Figure 10–18. The sentence is "He will list some words in that book," and there are several places in the sentence where the sounds do not have the form that would be observed if the individual words were spoken clearly. For example, in the sequence *list some*, there is a long fricative interval between 520 and 720 ms, but no evidence that a closure occurred for the stop consonant /t/. Or, in the word *that*, the initial consonant /ð/ (occurring at around 1300 ms) is produced as a nasal consonant rather

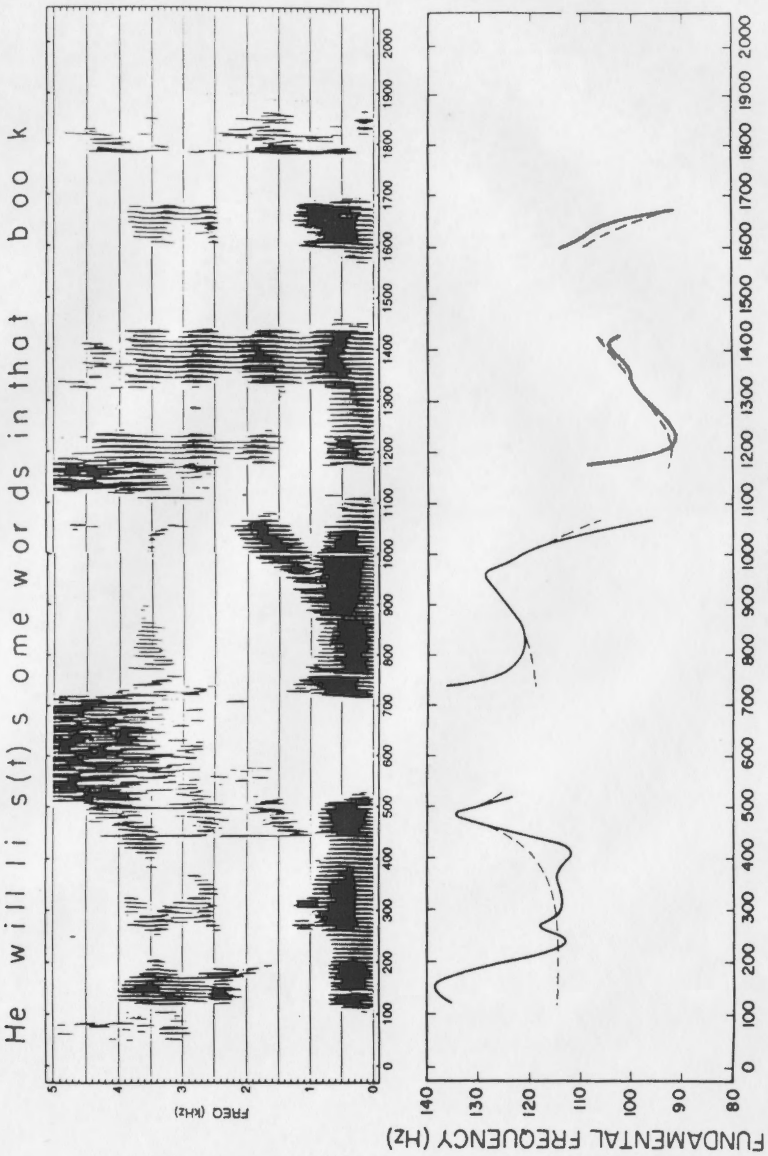

Figure 10–18. The upper panel is a spectrogram of the sentence: *He will list some words in that book*. The spectrogram illustrates some modifications of speech sounds when words are spoken in a sentence context. The solid line in the lower panel gives the contour of fundamental frequency versus time for the same sentence. The dashed line is an estimate of the shape of the fundamental frequency contour if local perturbations due to individual vowels and consonants were removed from the original contour. This dashed contour is intended to show which words in the sentence are produced with prominence and to mark the boundaries of the utterance. See text

than as the expected voiced fricative consonant. The kinds of modifications that can occur in running speech are constrained in particular ways, and native listeners appear to take these modifications into account without much difficulty.

The spectrogram in Figure 10–18 also illustrates the range of durations of phonetic units that can occur between speech sounds, especially when the sounds occur in longer utterances. For example, some syllables have greatly reduced prominence, and the durations of the vowels in these syllables are considerably shorter than the durations of vowels in syllables with greater prominence. In Figure 10–18, the vowels in *some* (740 ms) and *in* (1200 ms) undergo this kind of reduction. Other vowels receive special prominence in the sentence (i.e., the vowel in *words*), and consequently have a greater duration. Linguists refer to these changes in prominence as changes in syllable **stress**. In addition to the influence of the prominence of a syllable, several other factors contribute to the durations of the vowels and consonants. For example, some vowels are intrinsically longer than other vowels, and vowels at the end of a phrase or a sentence tend to be longer than vowels in the middle of the utterance.

The durations of vowels, then, as well as the durations of consonant units, provide information to a listener as to which syllables are more prominent than other syllables and also help to delineate the boundaries of phrases or sentences.

A speaker also marks prominences and boundaries by causing variations in the frequency of vocal fold vibration during the vowels in an utterance. The **fundamental frequency** (F_0) for a vowel can also undergo changes due to modifications of the vowel quality and to attributes of the consonant that precedes the vowel. The F_0 contour for the sentence "He will list some words in that book" is shown below the spectrogram in Figure 10–18. Some of the fluctuations in F_0 are a consequence of local effects of individual vowels and consonants. For example, the high F_0 on the first vowel is probably because the vowel /i/ is known to raise F_0 relative to its value for a vowel with a lower tongue-body position. Likewise, certain consonants that are adjacent to vowels can cause a local increase or decrease in the fundamental frequency near the vowel boundaries. An example is the increased F_0 at the beginning of the word *some* (at about 730 ms), caused by the voiceless consonant /s/. The main variations in F_0, however, are a consequence of the speaker's plan to place prominences on certain syllables, as well as to mark the boundaries of the utterance. In this example, the dashed line gives an estimate of the component of the fundamental frequency that is intended to mark the prominences and boundaries, with local effects due to particular vowels and consonants subtracted out. There are prominences on the words *list, words*, and *book*, with the final fall in fundamental frequency indicating the end of the utterance.

✦ SPEAKER DIFFERENCES AND SPEECH STYLES

We have focused up to now on the acoustic properties that distinguish one speech sound from another. It is well known, however, that the speech of a particular talker has individual characteristics that distinguish him or her from other speakers independent of the attributes that identify which sounds and words were spoken. We have seen, for example, that the frequency of vocal fold vibration for adult female speakers is higher, on the average, than it is for adult male speakers, and the formant frequencies for vowels spoken by women are, on the average, higher than for vowels produced by men. Speakers can differ from each other, however, in many other ways that are reflected in the acoustic patterns of their utterances. For example, people have different voice qualities, either because their laryngeal structures are different or because they choose to adjust the configuration of their vocal folds in different ways. These different voice qualities can be a consequence of different waveforms of glottal vibration, turbulence noise generation at the glottis, or irregular vibration of the vocal folds. Different talkers also can produce certain sounds or sequences of sounds, such as nasal consonants or particular types of vowels, in unique ways.

The acoustic characteristics of a particular speaker also may vary depending on the speaker's emotional state or the speaking situation. For example, the range of fundamental frequency, the glottal waveform, and the temporal pattern of the speech can be influenced by the talker's emotional or physiological state. There are a variety of disorders of speech production that can have a marked influence on the speech patterns of a talker.

To illustrate the kinds of differences that can occur between speakers, we show in Figure 10–19 spectra of the same vowel (in the word *but*) produced by two different female speakers. The first formant frequency is about the same for the two speakers (approximately 900 Hz), and the second formant frequencies are slightly different. Substantial differences between the speakers can be seen in the amplitudes of the first two harmonics (both of which are below 500 Hz) in relation to the amplitude of the first formant peak around 900 Hz. Also the first formant peak is much more prominent for speaker 2 (in the right panel) than for speaker 1. The spectrum for speaker 1 shows regularly spaced harmonics only up to about 1500 Hz, with an irregular spectrum shape at higher frequencies indicat-

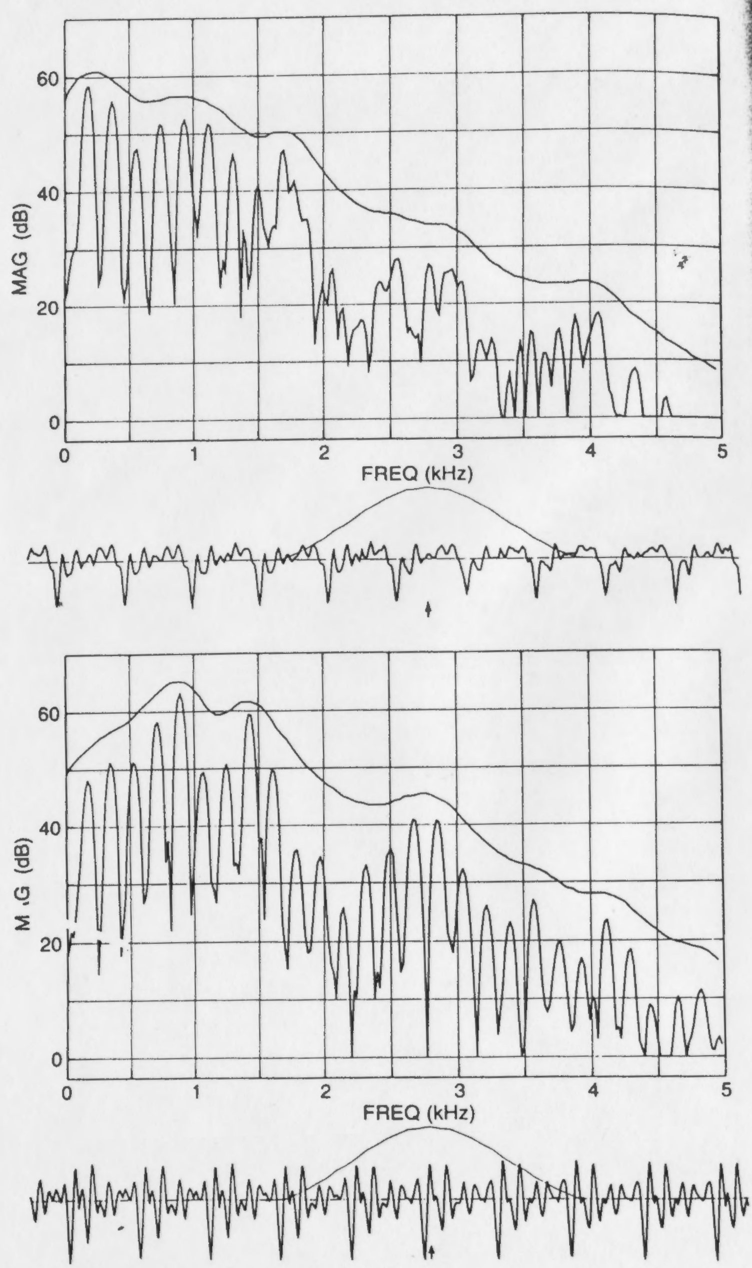

Figure 10-19. Spectra of the vowel in the word *but* produced by two adult female speakers. The spectra illustrate the differences that can occur between speakers. See text.

ing the presence of turbulence noise. For speaker 2, on the other hand, there are uniformly spaced harmonics throughout the frequency range up to 5000 Hz. Listeners hear the words with these two vowels as the same word *but*, but one voice (speaker 1) is clearly judged to be more breathy than the other. These differences are related to differences in the glottal sound source.

Figure 10-20 illustrates a more severe individual difference: a comparison of spectrograms of the production of the word *knot* by a normal speaker and by a speaker with a disorder in the control and coordination of the muscles used in speaking, known as dysarthria. One obvious difference is in the duration of the utterance, with the vowel and the final consonant closure being much longer for the dysarthric speaker. More detailed examination of the spectrograms shows that the dysarthric speaker exhibits some instability in the vowel and has difficulties in forming the final consonant. These kinds of quantitative acoustic analyses can have considerable value in diagnosing deviations from normal speech patterns, and in prescribing measures that might be taken for remediation of speech disorders. The differences seen in Figure 10-20 reflect temporal and spatial adjustments in the deployment of articulators in the vocal tract. These adjustments in the acoustic filter change the acoustic output of the vocal tract.

✦ SUMMARY AND CONCLUSIONS

The words that make up an utterance are represented in a speaker's memory as sequences of sounds or segments. Each of these seg-

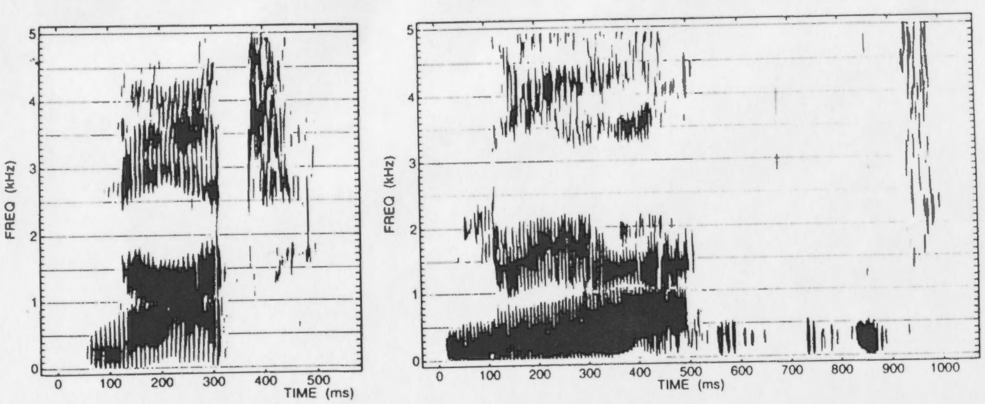

Figure 10-20. Spectrograms of the word *knot* produced by a normal speaker (*left panel*) and by a speaker with a disorder of muscle control (*right panel*).

ments contains instructions that specify how to move and to shape various articulators to produce the appropriate sound pattern. Movements of the articulators must be coordinated to perform two functions. One function is to generate sources of sound by modulating the flow of air in the vocal tract, either at the vocal folds or at a place in the upper airway where a constriction gives rise to turbulence in the airflow. The second function is to adjust the shape of the airway to filter the sources. Taken together, these two functions form the basis for the **source-filter theory of speech production**.

Descriptions of this acoustic process are most conveniently formulated in terms of source spectra and filtering mechanisms that modify these spectra. Speech sounds can be classified according to the kinds of sources that are used to generate the sounds and the nature of the filtering of these sources. A basic classification is in terms of vowels and consonants. Other categories are based on which articulator is used to form a constriction, how this articulator is positioned, and how the larynx is adjusted.

When an utterance consisting of a sequence of words is produced, a speaker can manipulate the durations of speech sounds and the frequency of vocal fold vibration to mark certain syllables as more prominent than others and to indicate boundaries between groups of words (e.g., at phrase and sentence boundaries). In casual speech, certain attributes of some speech sounds may be modified relative to their properties when they are produced slowly and clearly.

Speakers have particular characteristics that distinguish their speech from that of other speakers. Certain acoustic attributes of the speech of a given individual may vary depending on the emotional state of the speaker or the speech style that the speaker chooses to use in a particular situation. Acoustic analysis can also be used to interpret deviations from normal speech production.

✦ REFERENCES

Baer, T. (1975). *Investigation of phonation using excised larynxes.* Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Perkell, J. S. (1969). *Physiology of speech production: Results and implications of a quantitative cineradiographic study.* Cambridge, MA: MIT Press.

✦ SUGGESTED READINGS

Clark, J., & Yallop, C. (1990). *An introduction to phonetics and phonology.* Cambridge, MA: Blackwell.

Denes, P., & Pinson, E. (1973). *The speech chain.* Garden City, NY: Anchor Books.

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

Flanagan, J. L. (1972). *Speech analysis, synthesis, and perception.* Berlin: Springer-Verlag.

Kent, R. D., & Read, C. (1992). *The acoustic analysis of speech.* San Diego: Singular Publishing Group.

Kent, R. D., Atal, B. S., & Miller, J. L. (Eds.). (1991). *Papers in speech communication: Speech production.* Woodbury, NY: Acoustical Society of America.

Ladefoged, P. (1982). *A course in phonetics.* New York: Harcourt Brace Jovanovich.

Pickett, J. M. (1980). *The sounds of speech communication.* Baltimore: University Park Press.

Rosen, S., & Howell, P. (1991). *Signals and systems for speech and hearing.* San Diego: Academic Press.

✦ GLOSSARY

Affricate consonants: obstruent consonants produced with characteristics of both stop consonants and fricative consonants. Examples are the first and last sounds in the words *church* and *judge*.

Aperiodic: waveforms that are nonrepetitive. Waveforms that are not periodic.

Articulators: movable structures within the vocal tract or airway extending from the larynx to the mouth opening.

Aspiration: noise generated in the vicinity of the glottis.

Complex periodic waveform: a waveform constructed of two or more sinusoids whose frequencies are multiples of the frequency of the waveform.

Consonants: speech sounds produced by creating a major constriction in some region of the vocal tract.

Continuant: a consonant produced without complete closure of the vocal tract.

Diphthong: a vocalic glide within a single syllable (as in *bide* or *shout*).

Filtering: frequency-selective modifications of a sound source resulting from the resonating characteristics of the vocal tract.

Formants: peaks in the acoustic spectrum resulting from resonances within the vocal tract during speech production.

Formant transitions: time varying changes in resonant frequencies resulting from changing positions of articulators within the vocal tract during speech production.

Frequency: the number of complete cycles of a waveform per second. Frequency is measured in Hertz (Hz).

Fricative consonant: a continuant consonant produced by creating a narrow constriction within the vocal tract and supplying sufficient air flow to cause a continuing turbulence noise source.

Fundamental: the sinusoidal component of a complex wave whose frequency is equal to the repetition frequency of the waveform. The fundamental frequency is sometimes called the first harmonic.

Fundamental frequency: frequency of vocal fold vibration.

Glottis: the air channel opening between the vocal folds (in the larynx) through which air escapes during voice production, and through which air passes during normal respiration.

Hard palate: bony roof of the front of the mouth. Located anterior to the soft palate, which forms the back portion of the roof of the mouth.

Harmonics: whole number multiples of the fundamental frequency.

Larynx: the cartilaginous structure, located above the trachea and below the pharynx, that houses the vocal folds. The larynx is used to generate voiced sound and whisper during speech production.

Obstruent consonant: a consonant produced with a narrowing of the vocal tract that causes pressure to be built up behind the constriction.

Periodic: a pattern of sound pressure that is repetitive (repeats itself in the same amount of time).

Resonance: peaks in the transfer function of the vocal tract during acoustic filtering of the sound source. Resonance peaks are related to the speed of sound and the dimensions of the vocal tract.

Segment: an identifiable sound within a sequence of sounds that form a word.

Soft palate: the muscular portion of the roof of the mouth (also called the velum). Located posterior to the hard palate; forms the back portion of the roof of the mouth.

Sonorant consonant: a consonant produced with a narrowing of the vocal tract, but with no pressure build-up behind the narrowing.

Sound source: the generation of sound energy somewhere within the vocal tract, usually by narrowing of the airway.

Sound spectrum: a graph displaying frequency on the horizontal axis and amplitude on the vertical axis.

Source-filter theory of speech production: theory that describes the production of speech sounds as the generation of sound sources and the filtering of these sources by the vocal tract.

Spectrogram: a three-dimensional graph displaying frequency versus time, with spectrum amplitude represented by the degree of blackness in the pattern.

Stop consonant: a consonant sound produced as a result of a complete closure formed by an articulator in the vocal tract.

Stress: changes in the prominence of a syllable as a result of changes in fundamental frequency, intensity, and duration of the syllable.

Turbulence noise source: sound produced with a continuous spectrum. Turbulence noise sources result from the random variation of air pressure created when air particles pass through a narrow constriction at high velocity.

Velopharyngeal opening: the opening between the posterior portion of the oral cavity (actually oro-pharynx) and the nasal cavity.

Velum: soft palate, the back portion of the roof of the mouth.

Vocal folds: muscles and connective tissue located within the larynx that are used to produce voice (through vocal fold vibration). The two vocal folds are attached anteriorly to the thyroid cartilage and posteriorly to the arytenoid cartilages.

Vocal tract: the tortuously shaped airway extending from the larynx to the lips. The vocal tract can include the nasal airway during the production of nasal sounds.

Voice: sound produced by the larynx, usually through vibration of the vocal folds.

Voice onset time: the time interval between the release of pressure, built up as a result of vocal tract constriction during stop consonant production, and the onset of vocal fold vibration (voicing).

Vowel sounds: sounds produced when the vocal tract is relatively open (unconstricted).

Waveform: a graph of sound pressure versus time.