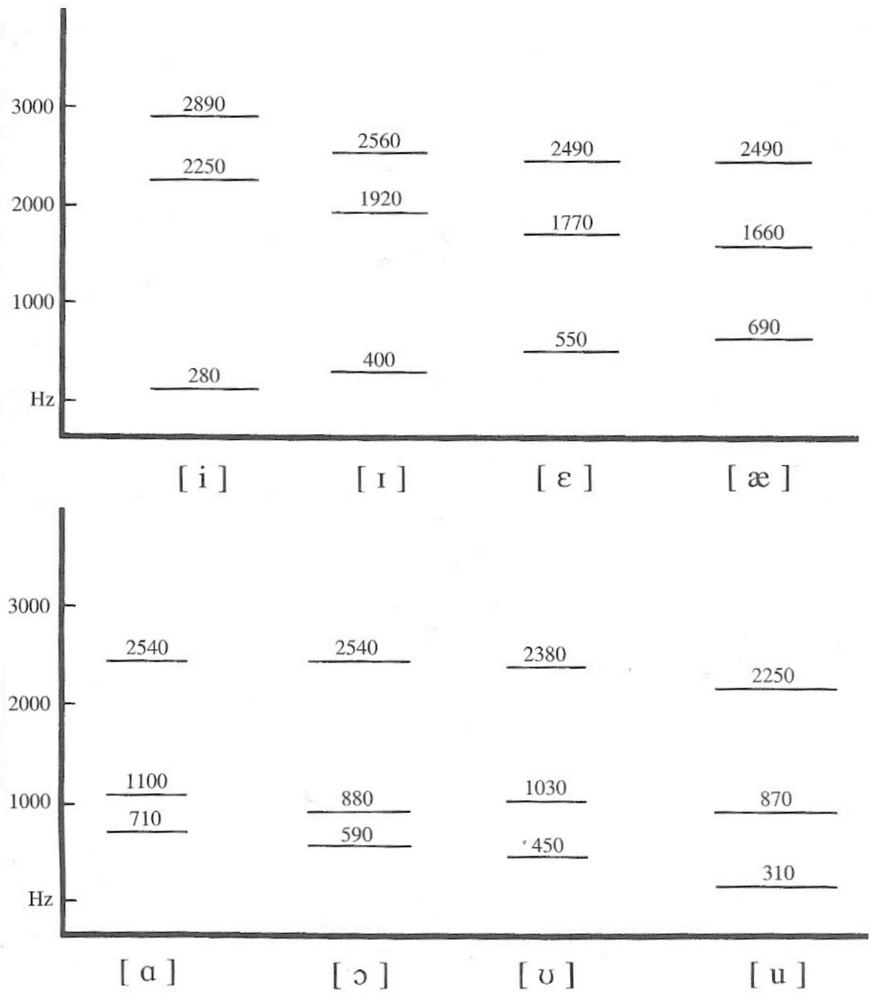
frequency of F2; one involves a very difficult constriction near the glottis, but without tongue root constriction (which is near the first V in the F2 resonance wave). The other involves constriction with the tongue against the roof of the mouth. This is the most common maneuver used to raise the F2 frequency.

ACOUSTIC ANALYSIS

Phonetic scientists like to describe vowels in terms of numbers. It is possible to analyze sounds so that we can measure the actual frequencies of the formants. We can then represent them graphically, as in Figure 8.3. This figure gives the average of a number of authorities' values of the frequencies of the first three formants in eight American English vowels. Try to see how your own vowels compare with these.

Figure 8.3 The frequencies of the first three formants in eight American English vowels.



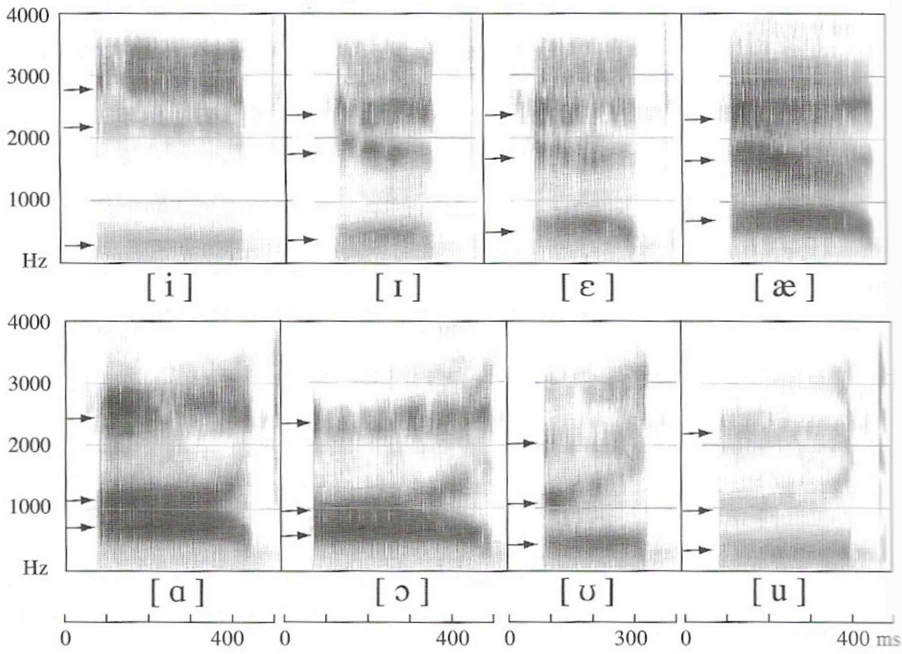
Do you have a much larger jump in the frequency of the second formant (which you hear when whispering) between [ɛ] and [æ] as compared with [ɪ] and [ɛ]? Do you distinguish between *hod* and *hawed* in terms of their formant frequencies?

There are computer programs that can analyze sounds and show their components. The display produced is called a **spectrogram**. We have seen spectrograms in prior chapters without much discussion of how to interpret them. Now is the time for a little more detail. In spectrograms, time runs from left to right, the frequency of the components is shown on the vertical scale, and the intensity of each component is shown by the degree of darkness. It is thus a display that shows, roughly speaking, dark bands for concentrations of energy at particular frequencies—showing the source and filter characteristics of speech. There are several free computer programs on the Web that can be used to make spectrograms. One of the best is WaveSurfer, from the Centre for Speech Technology (CTT) at KTH in Stockholm, Sweden. It is included (with permission) on the top level of the CD. You can open any of the sounds that you listen to on the CD and make spectrograms. If your computer has a built-in microphone, try recording your pronunciation of *heed, hid, head, had* and making spectrograms of them.

Figure 8.4 is a set of spectrograms of an American English speaker saying the words *heed, hid, head, had, hod, hawed, hood, who'd.* Because the higher

Figure 8.4 A spectrogram of the words *heed, hid, head, had, hod, hawed, hood, who'd* as spoken by a male speaker of American English. The locations of the first three formants are shown by arrows.

CD 8.3



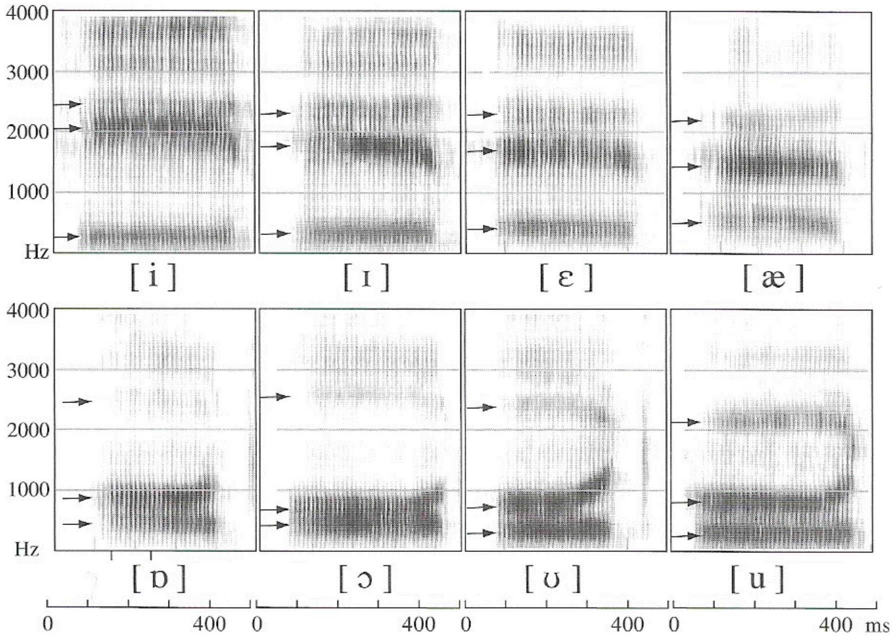
frequencies of the human voice have less energy, the higher frequencies have been given added emphasis. If they had not been boosted in this way, the higher formants would not have been visible. The time scale along the bottom of the picture shows intervals of 100 ms, so you can see that these words differ in length. The words were actually said one after another, but they have been put in separate frames as there was no point in showing the blank spaces between them. The vertical scale goes up to 4000 Hz, which is sufficient to show the component frequencies of vowels. Because the formants have greater relative intensity, shown by the darkness of the image, they can be seen as dark horizontal bars. The locations of the first three formants in each vowel are indicated by arrows.

There is a great deal of similarity between Figures 8.3 and 8.4. Figure 8.3 is like a schematic spectrogram of the isolated vowels. Figure 8.4 differs in that it represents a particular American English speaker rather than the mean of a number of speakers of American English. It also shows the effects of the consonant at the end of the word (which we will discuss later), and the slightly diphthongal character of some of the vowels. Note, for example, that the vowel [ɪ] starts with a higher second formant, and that the vowel [ʊ] has a large upward movement of the second formant. There is also a small downward movement of the second formant during [æ], indicating diphthongization of this vowel. In addition, there are some extra horizontal bars corresponding to higher formants that are not linguistically significant. The exact position of the higher formants varies a great deal from speaker to speaker. They are not uniquely determined for each speaker, but they certainly are indicative of a person's voice quality.

Figure 8.5 shows spectrograms of Peter Ladefoged's form of British English. It is similar to Figure 8.4, but not exactly the same because of the differences in accent and other individual differences. His head was larger than that of the American English speaker, so all his formants were slightly lower. Also, his vowels were less diphthongal—they had longer steady states. CD 8.4

Whenever the vocal folds are vibrating, there are regularly spaced vertical lines, close together, on the spectrogram. During a vowel, the vertical lines are visible throughout a large part of the spectrogram. Each vertical line in the vowels is the result of the momentary increase of acoustic energy due to a single movement of the vocal folds. We have seen that it is possible to observe the pulses in a record of the waveform and from this to calculate the pitch. It is equally possible to measure the pitch from observations of the vertical striations on spectrograms. When they are close together, the pitch is higher than when they are farther apart. At the bottom left of Figure 8.5, below the baseline but just above the symbol for [ɒ], there are two small lines, 100 ms apart. Within this tenth of a second, you can see that there are between eight and nine vertical striations in the vowel formants. The vocal folds must have been vibrating at about 85 Hz. This is not the best way of using spectrograms to determine the pitch. As we will see, it is possible to make another kind of spectrographic record that gives a better picture of the variations in pitch.

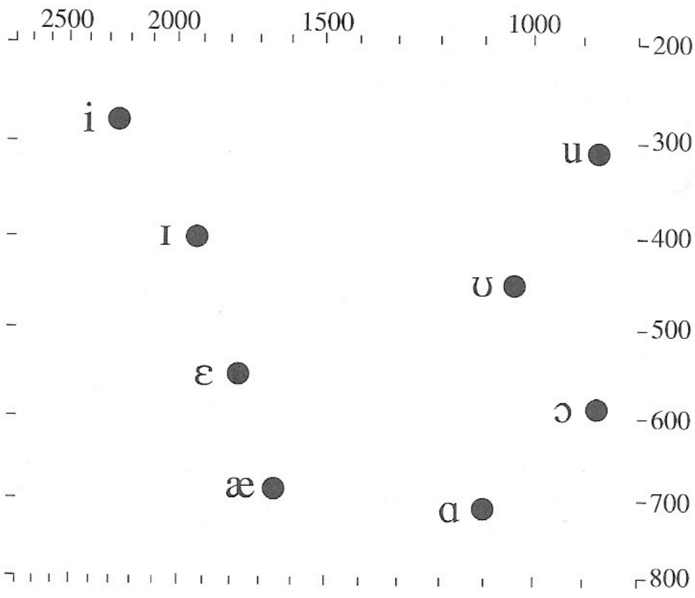
Figure 8.5 A spectrogram of the words *heed, hid, head, had, hod, hawed, hood, who'd* as spoken in a British accent. The locations of the first three formants are shown by arrows.



The traditional articulatory descriptions of vowels are related to the formant frequencies. We can see that the first formant frequency (indicated by the lowest of the three arrows in the frame for each vowel) increases as the speaker moves from the high vowel in *heed* to the low vowel in *had*. In these four vowels, the first formant frequency goes up as the vowel height goes down, both for the American English speaker in Figure 8.4 and for Peter Ladefoged in Figure 8.5. In the four vowels in the bottom rows of Figures 8.4 and 8.5, the first formant frequency decreases as the speaker goes from the low vowel in *hod* to the high vowel in *who'd*. Again in these vowels, the first formant frequency is inversely related to vowel height. We can also see that the second formant frequency is much higher for the front vowels in the top row than it is for the back vowels in the bottom row in each figure. But the correlation between the second formant frequency and the degree of backness of a vowel is not as good as that between the first formant frequency and the vowel height. The second formant frequency is considerably affected by the degree of lip rounding as well as by vowel height.

We can see some of the relationships between traditional articulatory descriptions and formants when we plot the formant frequencies given in Figure 8.3 along axes as shown in Figure 8.6. Because the formant frequencies

Figure 8.6 A formant chart showing the frequency of the first formant on the ordinate (the vertical axis) plotted against the second formant on the abscissa (the horizontal axis) for eight American English vowels. The scales are marked in Hz, arranged at Bark scale intervals.



are inversely related to the traditional articulatory parameters, the axes have been placed so that zero frequency would be at the top right corner of the figure rather than at the bottom left corner, as is more usual in graphical representations. In addition, the frequencies have been arranged in accordance with the Bark scale, in which perceptually equal intervals of pitch are represented as equal distances along the scale. As a further refinement, because the second formant is not as prominent as the first formant (which, on average, has 80 percent of the energy in a vowel), the second formant scale is not as expanded as the first formant scale. (Remember that in Figures 8.4 and 8.5, and in all the spectrograms in this book, the darkness scale does not correspond directly to the acoustic intensity of each sound. The higher frequencies have been given added emphasis to make them more visible.)

On a plot of formant frequencies, [i] and [u] appear at the top left and right of the graph, and [æ] and [ɑ] at the bottom, with all the other vowels in between. Consequently, this arrangement allows us to represent vowels in the way that we have become accustomed to seeing them in traditional articulatory descriptions.

In the preceding paragraphs, we have been careful to refer to the correlation between formant frequencies and the *traditional* articulatory descriptions. This is because, as we noted in Chapter 1, traditional articulatory descriptions are not entirely satisfactory. They are often not in accord with the actual articulatory facts. For well over a hundred years, phoneticians have been describing vowels in terms such as *high* versus *low* and *front* versus *back*. There is no doubt that these terms are appropriate for describing the relationships between different vowel qualities, but to some extent phoneticians have been using these terms as labels to specify acoustic dimensions rather than as descriptions of actual tongue positions. As G. Oscar Russell, one of the pioneers in x-ray studies of vowels, said, "Phoneticians are thinking in terms of acoustic fact, and using physiological fantasy to express the idea."

There is no doubt that the traditional description of vowel "height" is more closely related to the first formant frequency than to the height of the tongue. The so-called front–back dimension has a more complex relationship to the formant frequencies. As we have noted, the second formant is affected by both backness and lip rounding. We can eliminate some of the effects of lip rounding by considering the second formant in relation to the first. The degree of backness is best related to the difference between the first and the second formant frequencies. The closer they are together, the more "back" a vowel sounds.

Formant charts are commonly used to represent vowel qualities. To consolidate acoustic notions about vowels, you should now try to represent the vowels in Figures 8.4 and 8.5 in terms of a formant chart. We have provided arrows that mark what we take to be the formants that characterize these vowels. Measure these frequencies in terms of the scale on the left of each figure. Make a table listing the first and second formant frequencies and plot the vowels. A blank chart is provided as a PDF file on the CD.

CD 8.5

ACOUSTICS OF CONSONANTS

The acoustic structure of consonants is usually more complicated than that of vowels. In many cases, a consonant can be said to be a particular way of beginning or ending a vowel, and during the consonant articulation itself, there is no distinguishing feature. Thus, there is virtually no difference in the sounds during the actual closures of [b, d, g], and absolutely none during the closures of [p, t, k], for at these moments there is only silence.

Each of the stop sounds conveys its quality by its effect on the adjacent vowel. We have seen that during a vowel such as [ɛ] there will be formants corresponding to the particular shape of the vocal tract. These formants will be present as the lips open in a syllable such as [bɛ]. They will have frequencies corresponding to the particular shape that occurs at the moment that the lips come apart. As the lips come farther apart and the vocal tract shape changes, the formants will move correspondingly. As we saw in the section on perturbation theory, and Figure 8.2, closure of the lips causes a lowering of all the formants. Consequently, the syllable [bɛ] will begin with the formants in a lower position and will be distinguished by their rapidly rising to the positions for [ɛ]. Similarly, in the syllable [ɛb], the formants in [ɛ] will descend as the lip closure is formed.

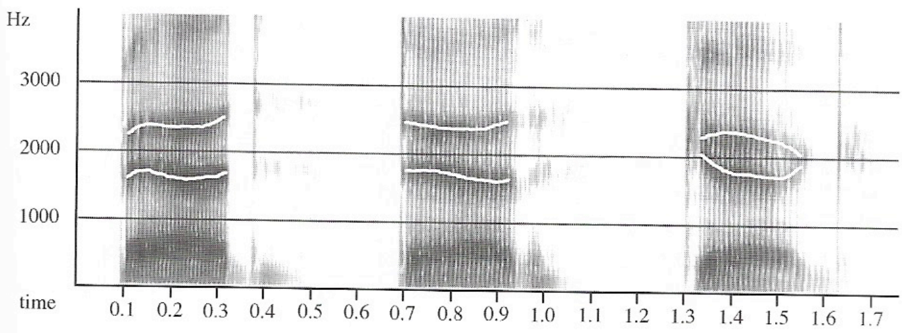
Whenever a stop is formed or released, there will be a particular shape of the vocal tract that will be characterized by particular formant frequencies.

When you say *bib* or *bab*, for example, the tongue will be in the position for the vowel even when the lips are closed at the beginning of the word. The formant frequencies at the moment of release will be determined by the shape of the vocal tract as a whole, and hence will vary according to the vowel. The apparent point of origin of the formant for each place of articulation is called the **locus** of that place of articulation. The point of origin of the formants will depend on the adjacent vowels. This is because the position of that part of the tongue that is not involved in the formation of a consonant closure will be largely that of the adjacent vowel.

Figure 8.7 shows spectrograms of the words *bed*, *dead*, and the nonword [gɛg] spoken by an American English speaker. You can see the faint voicing striations near the baseline for each of the final stops [d, d, g]. Evidence of voicing near the baseline during a consonant closure is called a **voice bar**. Note that this speaker, like many other speakers of English, has no voice bars in the initial "voiced" stops. CD 8.6

In all three words, the first formant rises from a low position. This is simply a mark of a stop closure and does not play a major part in distinguishing one place of articulation from another. What primarily distinguishes these three stops are the onsets and offsets of the second and third formants, which are traced with white lines. At the beginning of the word *bed*, the second and third formants have a lower frequency than they do at the beginning of the word *dead*. The second formant is noticeably rising for the initial [b] from a comparatively low locus. In the word *dead*, the second formant is fairly steady at the beginning and the third formant drops a little. In [gɛg], the second and third formants come close to each other at the margins of the vowel, where the [g] consonants have the most influence over the formant frequencies. It is almost as if the F2 and F3 were going to a common point. This coming together of the second and third formants, sometimes called a **velar pinch**, is very characteristic of velar consonants.

Figure 8.7 Spectrograms of the words *bed*, *dead*, and the nonword [gɛg] spoken by an American English speaker.

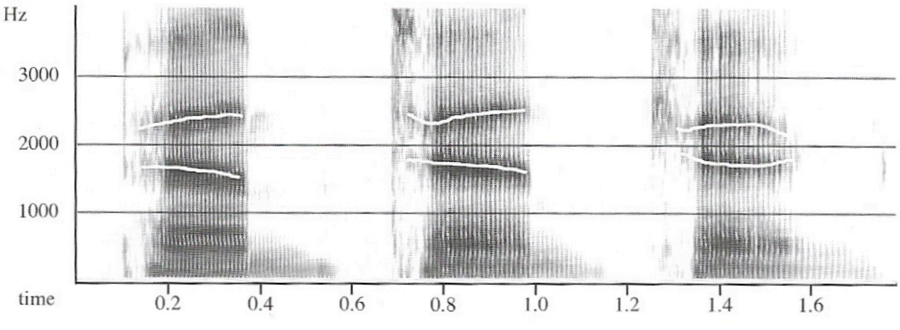


The corresponding voiceless stops, [p, t, k], are illustrated in Figure 8.8, in the forms [pʰɛm], [tʰɛn], and [kʰɛŋ], of which only *ten* is an English word. We chose these forms because the vowel [ɛ] is a particularly good environment for showing stop-consonant place-of-articulation differences. As with unaspirated stops, the release of the aspirated stops is marked by a sudden sharp spike corresponding to the onset of a burst of noise. After the release burst, there is a period of aspiration noise marked by absence of energy in F1 and absence of the regular vertical striations of voicing. The aspiration noise separates the burst from the voiced portion of the vowel. The burst for [p] has the lowest frequency. For both [t] and [k], the noise extends above the 4000 Hz shown in the spectrogram, as we will see in later figures. The highest frequencies are actually in the [t] burst rather than the [k]. If you whisper a sequence of consonants [t, t, t, k, k, k, p, p, p] in that order, [t, k, p], you can hear that the highest pitch is associated with [t], the next with [k], and the lowest with [p]. You can also hear that [t] is the loudest, [k] next loudest, and [p] the least loud. The intensity of the [p] burst is sometimes so low that there is hardly any evidence of a sharp spike in the spectrogram. Since the formant transitions after voiceless aspirated stops take place during the period of aspiration, they are not as apparent in Figure 8.8 as they are after the voiced stops in Figure 8.7. However, we have traced the centers of F2 and F3 in these spectrograms to help you see that the formants are also present in the aspiration noise. In addition, the transitions into the final nasals from the vowels before them are easily visible. The second and third formants are falling (slightly) before [m]; the second and third formants are almost level before [n]. Most distinctive of all, the second and third formants are coming together for the velar pinch before [ŋ].

CD 8.7

The nasal consonants [m, n, ŋ] are also illustrated in Figure 8.8. A clear mark of a nasal (or, as we will see, a lateral) consonant is an abrupt change in the spectrogram at the time of the formation of the articulatory closure. Each of the nasals has a formant structure similar to that of a vowel, except that the bands are fainter

Figure 8.8 Spectrograms of the forms [pʰɛm], [tʰɛn] (*ten*), and [kʰɛŋ].

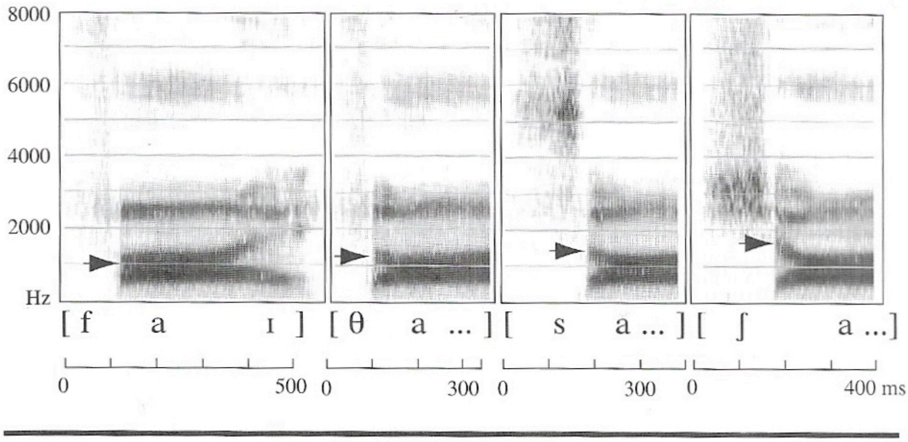


and are in particular frequency locations that depend on the characteristic resonances of the nasal cavities. In nasal consonants, there is usually a very low first formant centered at about 250 Hz. The location of the higher formants varies, but generally there is a large region above the first formant with no energy. This speaker has a second, rather faint, nasal formant around 2000 Hz. The difference between each of the nasals is often determinable from the different formant transitions that occur at the end of each vowel. There is a decrease in the second formant of the vowel before [m], and formants two and three are coming together for the velar pinch before the velar nasal at the end of [kʰɛŋ]. But the place cues are sometimes not very clear.

Figure 8.9 shows the words *fie, thigh, sigh, shy* illustrating the voiceless fricatives. The frequency scale for these spectrograms has been increased to 8000 Hz, as the highest frequencies in speech occur during fricatives. In [s] sounds, the random noise extends well beyond the upper limits of even this spectrogram. The spectrogram of the first word, *fie*, shows the diphthong that occurs in each of these words. The first and second formants in this diphthong start close together in the position for a low central vowel. They then move apart so that at the end of the diphthong they are in locations similar to those in [ɪ] in Figure 8.3. As the formant pattern for the diphthong is the same in *fie, thigh, sigh, shy*, only the first part has been shown for the last three words. **CD 8.8**

All these sounds have random energy distributed over a wide range of frequencies. In [f] and [θ], the pattern is much the same. What distinguishes these two words is the movement of the second formant into the following vowel, marked by arrows in the figure. There is very little movement in [f], but in [θ], the second formant starts at around 1200 Hz and moves down.

Figure 8.9 A spectrogram of *fie, thigh, sigh, shy*. The frequency scale goes up to 8000 Hz in this figure. The arrows mark the onsets of the second formant transitions. Only the first word is shown in full. The second part of the diphthong has been deleted for the other words.



Because the differences between these two sounds are so small, they are often confused in noisy settings, and they have fallen together as one sound in some accents of English, such as London Cockney, which does not distinguish between *fin* and *thin*.

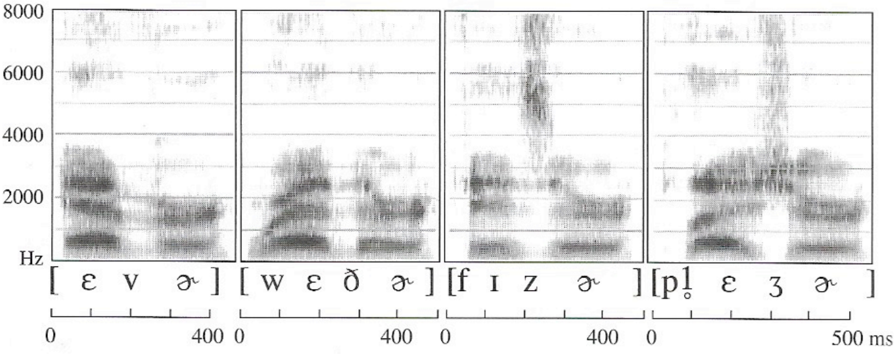
The noise in [s] is centered at a high frequency, between 5000 and 6000 Hz in Figure 8.9. In [ʃ], it is lower, extending down to about 2500 Hz. Since both [s] and [ʃ] have a comparatively large acoustic intensity, they produce darker patterns than [f] or [θ]. They are also marked by distinctive formant transitions. The apparent origin (the locus) of the second formant transition increases throughout the four words *fie, thigh, sigh, shy* so that in *shy*, it is in a position comparable to its location in the vowel [i] and falls considerably.

The voiced fricatives corresponding to [f, θ, s, ʃ] do not contrast at the beginnings of words. Accordingly, Figure 8.10 shows [v, ð, z, ʒ] between vowels. These voiced fricatives have patterns similar to their voiceless counterparts, but with the addition of the vertical striations indicative of voicing. The fricative component of [v] in *ever* is even fainter than the [f] in *face* and is really only visible at the start of the following vowel. The vertical striations due to voicing are apparent throughout the articulation. The same is true of [ð] in *whether*. As with their voiceless counterparts, [f, θ], it is the formants in the adjacent vowels that distinguish these words. In this figure, both these fricatives are preceded by [ɛ] and followed by [ɚ]. The second formants are much higher around [ð] than around [v].

CD 8.9

The fricative energy in the higher frequencies is very apparent in [z] and [ʒ]. There is a faint voice bar in [z], but in [ʒ], the voicing is hard to see. There are only a few vertical striations due to voicing in the 6000- to 8000-Hz range at the beginning of the fricative noise. The formant transition from [z] into the vowel [ɚ] is level, but that from [ʒ] falls considerably. This last word, *pleasure*, also enables us to see what happens when an aspirated stop such as [p] is followed by an approximant such as [l]. Most of the [l] is voiceless, audible only by the effect it has on the [p] burst and the aspiration noise.

Figure 8.10 A spectrogram of *ever, weather, fizzer, pleasure.*



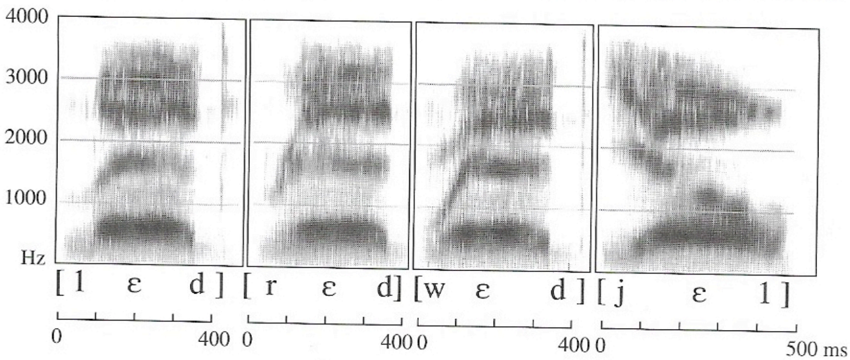
The last set of English consonants to consider are the lateral and central approximants, [l, r, w, j]. Figure 8.11 shows these sounds in the words *led, red, wed, yell*. All these voiced approximants have formants not unlike those of vowels. The initial lateral in the first word has formants with center frequencies of approximately 250, 1100, and 2400 (low intensity), which change abruptly in intensity at the beginning of the vowel. As we noted above, a marked change in the formant pattern is characteristic of voiced nasals and laterals. At the end of a word, as in *yell* in Figure 8.11, there may be a less marked change. A final lateral may have little or no central contact, making it not really a lateral but a back unrounded vowel. A formant in the neighborhood of 1100 or 1200 Hz is typical of most initial laterals for most speakers.

CD 8.10

The second word in Figure 8.11 illustrates the approximant [r] in *red*. (Remember that in the broad transcription of English used in this book, the symbol [r] is used for the approximant [ɹ].) The most obvious feature of this kind of [r] is the low frequency of the second and third formants. The third formant in particular has a very low frequency. In this example, its origin (above the symbol [r]) is around 1600 Hz. There is a great deal of similarity between *red* and the third word, *wed*, which is why young children sometimes have difficulty learning to distinguish them. The approximant [w] also starts with a low position of all three formants, but this time, it is the second formant that has the sharpest rise. The movements of the formants for [w] are like those in a movement away from a very short [u] vowel. Finally, the movements of the formants for [j], as in *yell* or *yes*, are like those in a movement away from a very short [i] vowel. Both [w] and [j] are appropriately called semivowels.

The vagueness of many of the remarks in the preceding paragraphs is meant to convey that the interpretation of sound spectrograms is often not all straightforward. The acoustic correlates of some articulatory features are summarized in Table 8.1. But in a book such as this, it is impossible to give a completely detailed account of the acoustics of speech. The descriptions that have been given should

Figure 8.11 A spectrogram of *led, red, wed, yell*.



Because the differences between these two sounds are so small, they are often confused in noisy settings, and they have fallen together as one sound in some accents of English, such as London Cockney, which does not distinguish between *fin* and *thin*.

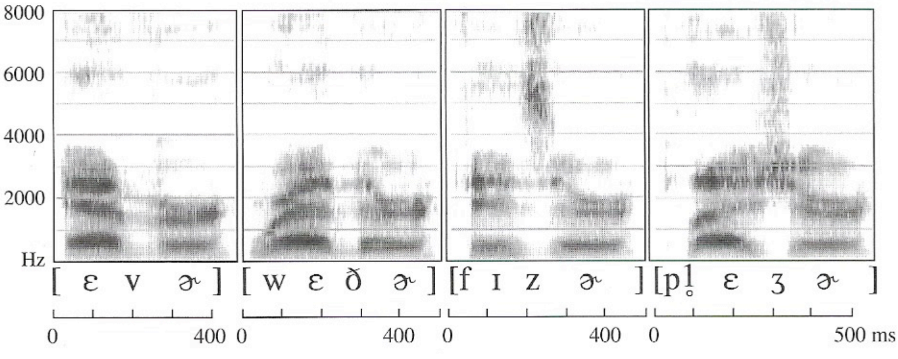
The noise in [s] is centered at a high frequency, between 5000 and 6000 Hz in Figure 8.9. In [ʃ], it is lower, extending down to about 2500 Hz. Since both [s] and [ʃ] have a comparatively large acoustic intensity, they produce darker patterns than [f] or [θ]. They are also marked by distinctive formant transitions. The apparent origin (the locus) of the second formant transition increases throughout the four words *fie*, *thigh*, *sigh*, *shy* so that in *shy*, it is in a position comparable to its location in the vowel [i] and falls considerably.

The voiced fricatives corresponding to [f, θ, s, ʃ] do not contrast at the beginnings of words. Accordingly, Figure 8.10 shows [v, ð, z, ʒ] between vowels. These voiced fricatives have patterns similar to their voiceless counterparts, but with the addition of the vertical striations indicative of voicing. The fricative component of [v] in *ever* is even fainter than the [f] in *face* and is really only visible at the start of the following vowel. The vertical striations due to voicing are apparent throughout the articulation. The same is true of [ð] in *whether*. As with their voiceless counterparts, [f, θ], it is the formants in the adjacent vowels that distinguish these words. In this figure, both these fricatives are preceded by [ɛ] and followed by [ɚ]. The second formants are much higher around [ð] than around [v].

CD 8.9

The fricative energy in the higher frequencies is very apparent in [z] and [ʒ]. There is a faint voice bar in [z], but in [ʒ], the voicing is hard to see. There are only a few vertical striations due to voicing in the 6000- to 8000-Hz range at the beginning of the fricative noise. The formant transition from [z] into the vowel [ɚ] is level, but that from [ʒ] falls considerably. This last word, *pleasure*, also enables us to see what happens when an aspirated stop such as [p] is followed by an approximant such as [l]. Most of the [l] is voiceless, audible only by the effect it has on the [p] burst and the aspiration noise.

Figure 8.10 A spectrogram of *ever, weather, fizzer, pleasure.*



The last set of English consonants to consider are the lateral and central approximants, [l, r, w, j]. Figure 8.11 shows these sounds in the words *led, red, wed, yell.* All these voiced approximants have formants not unlike those of vowels. The initial lateral in the first word has formants with center frequencies of approximately 250, 1100, and 2400 (low intensity), which change abruptly in intensity at the beginning of the vowel. As we noted above, a marked change in the formant pattern is characteristic of voiced nasals and laterals. At the end of a word, as in *yell* in Figure 8.11, there may be a less marked change. A final lateral may have little or no central contact, making it not really a lateral but a back unrounded vowel. A formant in the neighborhood of 1100 or 1200 Hz is typical of most initial laterals for most speakers. **CD 8.10**

The second word in Figure 8.11 illustrates the approximant [r] in *red.* (Remember that in the broad transcription of English used in this book, the symbol [r] is used for the approximant [ɹ].) The most obvious feature of this kind of [r] is the low frequency of the second and third formants. The third formant in particular has a very low frequency. In this example, its origin (above the symbol [r]) is around 1600 Hz. There is a great deal of similarity between *red* and the third word, *wed*, which is why young children sometimes have difficulty learning to distinguish them. The approximant [w] also starts with a low position of all three formants, but this time, it is the second formant that has the sharpest rise. The movements of the formants for [w] are like those in a movement away from a very short [u] vowel. Finally, the movements of the formants for [j], as in *yell* or *yes*, are like those in a movement away from a very short [i] vowel. Both [w] and [j] are appropriately called semivowels.

The vagueness of many of the remarks in the preceding paragraphs is meant to convey that the interpretation of sound spectrograms is often not all straightforward. The acoustic correlates of some articulatory features are summarized in Table 8.1. But in a book such as this, it is impossible to give a completely detailed account of the acoustics of speech. The descriptions that have been given should

Figure 8.11 A spectrogram of *led, red, wed, yell.*

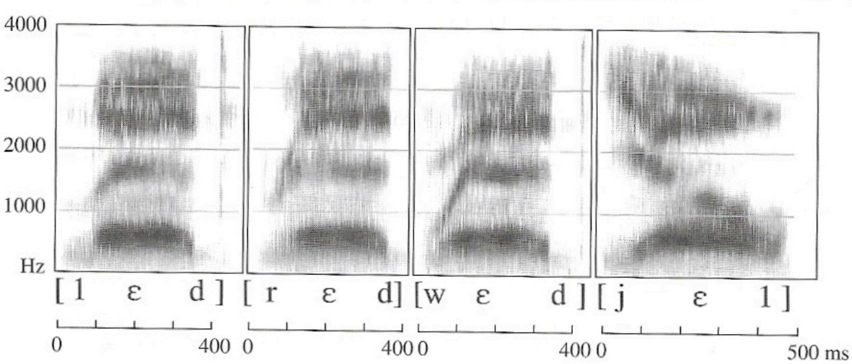


TABLE 8.1 Acoustic correlates of consonantal features. Note: These descriptions should be regarded only as rough guides. The actual acoustic correlates depend to a great extent on the particular combination of articulatory features in a sound and on the neighboring vowels.

Voiced	Vertical striations corresponding to the vibrations of the vocal folds.
Bilabial	Locus of both second and third formants comparatively low.
Alveolar	Locus of second formant about 1700–1800 Hz.
Velar	Usually high locus of the second formant. Common origin of second and third formant transitions.
Retroflex	General lowering of the third and fourth formants.
Stop	Gap in pattern, followed by burst of noise for voiceless stops or sharp beginning of formant structure for voiced stops.
Fricative	Random noise pattern, especially in higher frequency regions, but dependent on the place of articulation.
Nasal	Formant structure similar to that of vowels but with nasal formants at about 250, 2500, and 3250 Hz.
Lateral	Formant structure similar to that of vowels but with formants in the neighborhood of 250, 1200, and 2400 Hz. The higher formants are considerably reduced in intensity.
Approximant	Formant structure similar to that in vowels, usually changing.

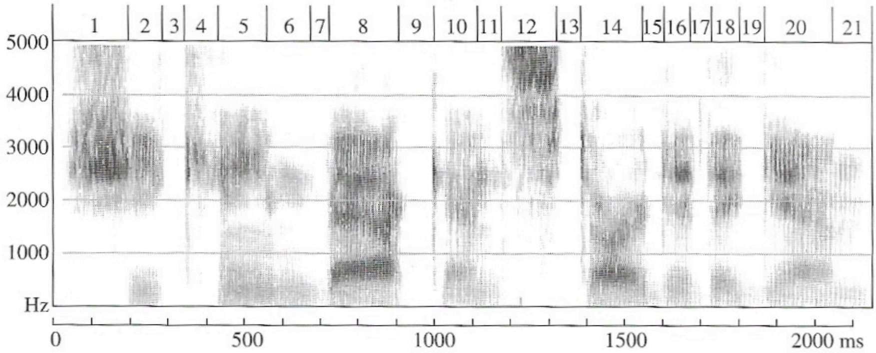
be regarded as rough guides rather than accounts of invariable structures that can always be seen in spectrograms. When any of the segments described above occurs in a different phonetic context, it may have a surprisingly different acoustic structure.

INTERPRETING SPECTROGRAMS

All the words illustrated in spectrograms so far were spoken in a fairly distinct way. In connected speech, as in the remainder of the spectrograms illustrating this chapter, many of the sounds are more difficult to distinguish. Before reading
CD 8.11 the next paragraph, transcribe the segments in Figure 8.12, given the information that the utterance was *She came back and started again*, as spoken by the speaker who produced the vowels in Figure 8.3.

Looking at the segments one at a time, we can see that the initial [ʃ] sound is similar to that in *shy* in Figure 8.9. The frequency scale is not as extended as that in Figure 8.9 (so that more attention can be paid to the vowel formants), but it is quite easy to see that this is [ʃ], not [s] as in segment (12), which has a higher frequency. The second segment, [i], has second and third formant frequencies that are a little lower than in this speaker's vowel in Figure 8.3. At the end of segment (2), the second and third formants come together for the velar stop that forms segment (3). This stop is followed by a burst of aspiration—marked as segment (4)—before the onset of the vowel. The vowel in *came*, (5), is a diphthong, [eɪ], with a faint additional formant around 1100 Hz, associated with the

Figure 8.12 A spectrogram of *She came back and started again.*



nasalization of the vowel. At the end of the bilabial nasal (6), there is a short [b] closure (7), in which the voicing is just barely visible. The upward transitions after the bilabial stop at the beginning of [æ] (8) in *back* are much more evident. There is no difficulty in seeing the coming together of the second and third formants before the velar stop [k] (9). There is only a short period of aspiration, not given a separate segment number, followed by a transition, the coming apart of the second and third formants, before a neutral vowel, [ə] (10). This is followed by an alveolar nasal [n] (11).

The [s] (12) in *started* is followed by a short [t] (13), which is only slightly aspirated (as is normal for [t] whenever it occurs after [s] in English). The falling second formant into the vowel, [ɑ] (14), is typical of the transition from [t] into [ɑ]. The falling of the third formant for the last part of segment (14) is associated with the *r*-coloring. Approximately the last half of the vowel is rhotacized. The stop in (15) has a voice bar and could be symbolized by [ɾ] in a narrow phonetic transcription. For many people, including this speaker, past-tense *-ed* forms after an alveolar stop have a fairly high second formant and a low first formant. The vowel in segment (16) is probably better as [ɪ] rather than [ə]. Segment (17), like (15), is a tap [ɾ]. The vowel in segment (18) is also [ɪ]; unstressed vowels before velar consonants are often [ɪ] rather than [ə]. The velar stop [g] in (19) is clearly marked by the coming together of the second and third formants in the vowels on either side of it. The final syllable in *again* has a fairly low vowel—formant one is about as high as it is in segment (8), the vowel [æ] in *back*. Segment (20) could be transcribed as [ɛ] or [æ]. Segment (21) is the final nasal [n].

Now you should try segmenting a more difficult utterance. Figure 8.13 shows a spectrogram of Peter Ladefoged saying, *I should have thought spectrograms were unreadable.* This phrase was spoken in a normal, but rapid, conversational style. This time, instead of marking the separate segments, we have simply placed evenly spaced lines above the spectrogram so that we can refer to

CD 8.12