8

Acoustic Phonetics

FORMANTS

In the first chapter of this book we discussed how speech sounds can differ in pitch, in loudness, and in quality. When discussing differences in quality we noted that the quality of a vowel depends on its overtone structure. Putting this another way, we can say that a vowel sound contains a number of different pitches simultaneously. There is the pitch at which it is actually spoken, and there are the various overtone pitches that give it its distinctive quality. We distinguish one vowel from another by the differences in these overtones. To review what was said earlier, we found that each vowel had three formants, three overtone pitches. The lowest, formant one, which we can symbolize as F1, could be heard most easily when the vowels were produced with a creaky voice. You can hear (in your own speech or in the recordings for Chapter 1 on the CD) that in some sense the pitch of the vowels in *heed, hid, head, had* goes up when these vowels are said with a creaky voice, one with no real pitch itself. The second formant in this series of vowels, F2, goes down in pitch, as can be heard more easily when these vowels are whispered. The third formant, F3, adds to quality distinctions, but there is no easy way of making it more evident.

CD 1.4

How do these formants arise? The answer is that the air in the vocal tract acts like the air in an organ pipe, or in a bottle. When you give it a tap it will vibrate. If you open your mouth, make a glottal stop, and flick a finger against your neck just to the side and below the jaw, you will hear a note, just as you would if you tapped on a bottle. If you tilt your head slightly backward so that the skin of the neck is stretched while you tap, you may be able to hear this sound somewhat better. Be careful to maintain a vowel position and not to raise the back of the tongue against the soft palate. If you check a complete set of vowel positions [i, ɪ, e, ɛ, æ, ɑ, ɔ, ʊ, u] with this technique, you should hear the pitch of the first formant going up for the first four vowels and down for the second four vowels.

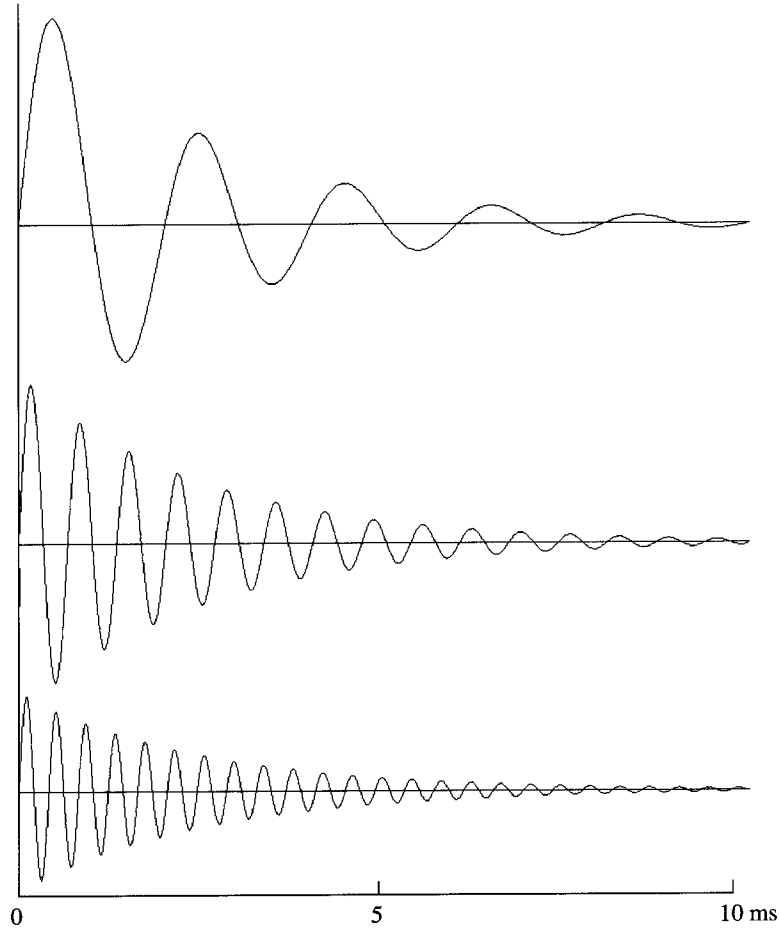
The formants that characterize different vowels are the result of the different shapes of the vocal tract. Any body of air, such as that in the vocal tract or that in a bottle, will vibrate in a way that depends on its size and shape. If you blow across the top of an empty bottle, you can produce a low-pitched note. If you partially fill the bottle with water so that the volume of air is smaller, you will be able to produce a note with a higher pitch. Smaller bodies of air are similar to smaller piano strings or smaller organ pipes in that they produce higher pitches. In the case of vowel sounds, the vocal tract has a complex shape so that the different bodies of air produce a number of overtones.

The air in the vocal tract is set in vibration by the action of the vocal folds. Every time the vocal folds open and close, there is a pulse of air from the lungs. These pulses act like sharp taps on the air in the vocal tract, setting the resonating cavities into vibration so that they produce a number of different frequencies, just as if you were tapping on a number of different bottles at the same time. Irrespective of the rate of vibration of the vocal folds, the air in the vocal tract will resonate at these frequencies as long as the position of the vocal organs remains the same. Because of the complex shape of the tract the air will vibrate in more than one way at once. It's as if the air in the back of the vocal tract might vibrate one way, producing a waveform like that at the top of of Figure 8.1, while the air in front of the tongue, a smaller cavity, might vibrate in another way, producing a waveform more like the second wave in the figure. A third mode of vibration of the air in the vocal tract might produce the third wave in Figure 8.1. What we actually hear would be the sum of these waveforms added together.

Look at the top waveform in Figure 8.1. As it is the product of a single tap on the vocal tract, it is a wave with decreasing amplitude (it gets smaller as time goes on). You can see from the time scale at the bottom of the figure that there are five peaks of air pressure within 10 ms, which is one-hundredth of a second. This corresponds to 500 peaks within one second. In other words, this is a 500-Hz wave, approximately the value of F1 in the vowel [ə]. The other two waves in the figure have higher frequencies—there are more peaks of air pressure within the 10 ms. They correspond to 1500- and 2500-Hz waves, the frequencies of F2 and F3 in [ə]. If the air in the vocal tract receives a single blow, it will produce sound waves something like the slightly stylized waveforms in Figure 8.1. If it receives multiple blows as air passes through the vibrating vocal folds, these waveforms will be produced repeatedly, adding a fundamental pitch to these overtone frequencies.

In this book we will not go into detail concerning the relationship between vocal tract shapes and the formant frequencies. That topic is treated more fully in another of my books, *Elements of Acoustic Phonetics* (Chicago, 1996). The relationship is actually much more complicated than the air in the back part of the vocal tract vibrating in one way and the air in other parts vibrating in another. Here we will just concentrate on the fact that in most voiced sounds three formants are produced every time the vocal folds vibrate. Note that the rate of

FIGURE 8.1 Three waves that might be produced by a single tap on the air in the vocal tract.



vibration of the air in the vocal tract is independent of the rate of vibration of the vocal folds. The vocal fold may vibrate faster or slower, giving the sound a higher or lower pitch, but the formant frequencies will remain the same as long as there are no changes in the shape of the vocal tract.

There is nothing particularly new about this way of analyzing vowel sounds. The general theory of formants was stated by the great German scientist Hermann Helmholtz about one hundred fifty years ago. Even earlier, in 1829, the English physicist Robert Willis had said, "A given vowel is merely the rapid repetition of its peculiar note." We would nowadays say that a vowel is the rapid repetition (corresponding to the vibrations of the vocal folds) of its peculiar two or three notes (corresponding to its formants). We can, in fact, go even further

and say that not only vowels but all voiced sounds are distinguishable from one another by their formant frequencies.

The notion that a vowel contains several different pitches at the same time is difficult to appreciate. One way of making it clearer is to build up a sentence from the component waves. The speech-synthesis demonstration on the CD shows how this can be done. You can listen to the components of the sentence *A bird in the hand is worth two in the bush* in a synthesized version of my voice. The first link below the table on the CD enables you to hear just the variations in the first formant, which sounds like a muffled version of the sentence. The vocal fold pulses have been produced at a steady rate, so that the "utterance" is on a monotone. What you hear as the changes in pitch are the changes in the overtones of this monotone "voice." These overtone pitch variations convey a great deal of the quality of the voiced sounds. The rhythm of the sentence is apparent because the overtone pitches occur only when the vocal folds would have been vibrating. The amplitude (loudness) of the first formant is turned up only at these times.


CD 8.1

The second link below the table on the CD does the same for the second formant. This time the equivalent of a series of monotone vocal fold pulses has been used to excite only the second formant. Again the variations of these overtone pitches convey much of the vowel quality. The same is not so true of the third formant by itself, which you can hear by playing the third link. This formant adds to the overall quality of the sound, but, in this sentence, it does not play a very significant role.


CD 8.1

The fourth link plays the sound of the three formants added together. With this, the sentence becomes highly intelligible. A slight improvement in quality occurs by adding some additional, fixed, formants, which you can hear by playing the fifth link. At this point in the synthesis of the sentence, everything is there except the bursts of noise associated with the releases of the stop consonants and the turbulent noises of the fricatives. Play this link again and note that, for example, the final [ʃ] in *bush* is not there.


CD 8.1

The sixth link enables you to hear the sounds of the bursts of noise and the turbulence of the fricatives by themselves. When they are added in the correct places, as they are for the seventh link, you can hear the entire sentence in a monotone. The last link adds the fundamental pitch, which varies as the glottal pulses recur at different intervals, so that the sentence is pronounced with a reasonable intonation.


CD 8.1

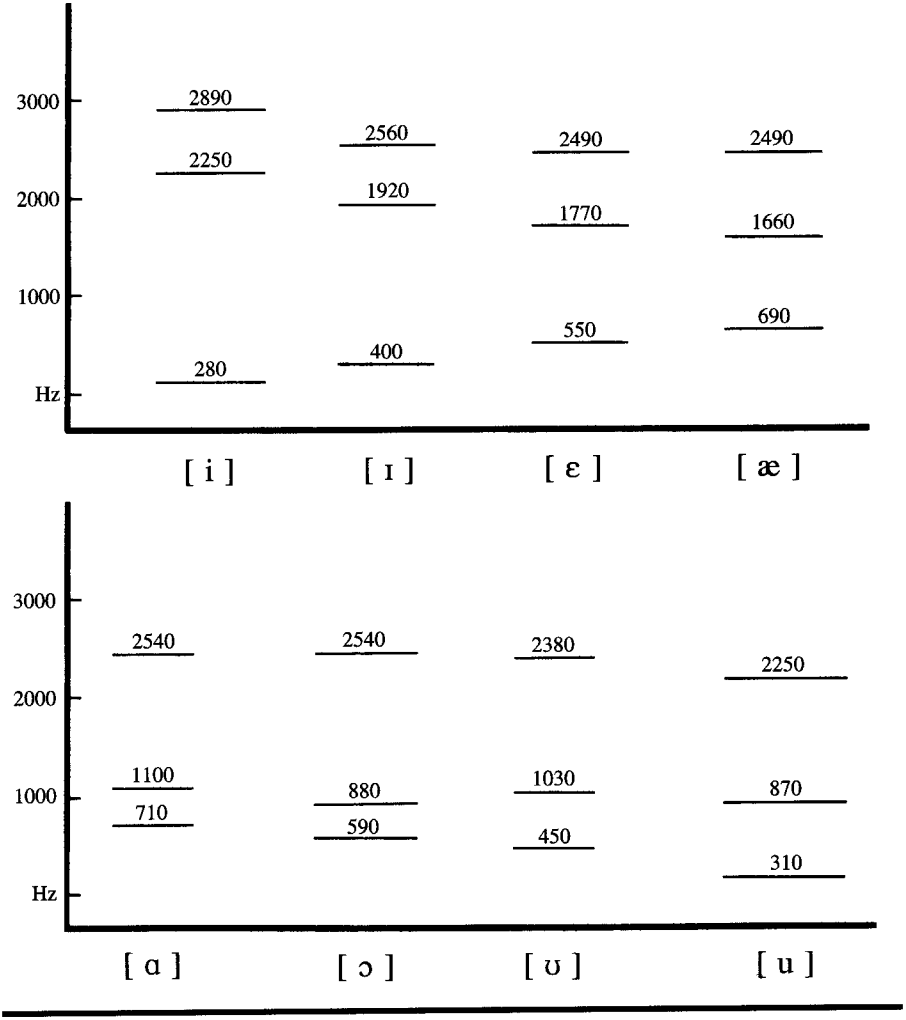
ACOUSTIC ANALYSIS

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 8.2. 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. Do you have a much larger jump in the pitch 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,** in which 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 each of the groups of overtone

FIGURE 8.2 The frequencies of the first three formants in eight American English vowels.

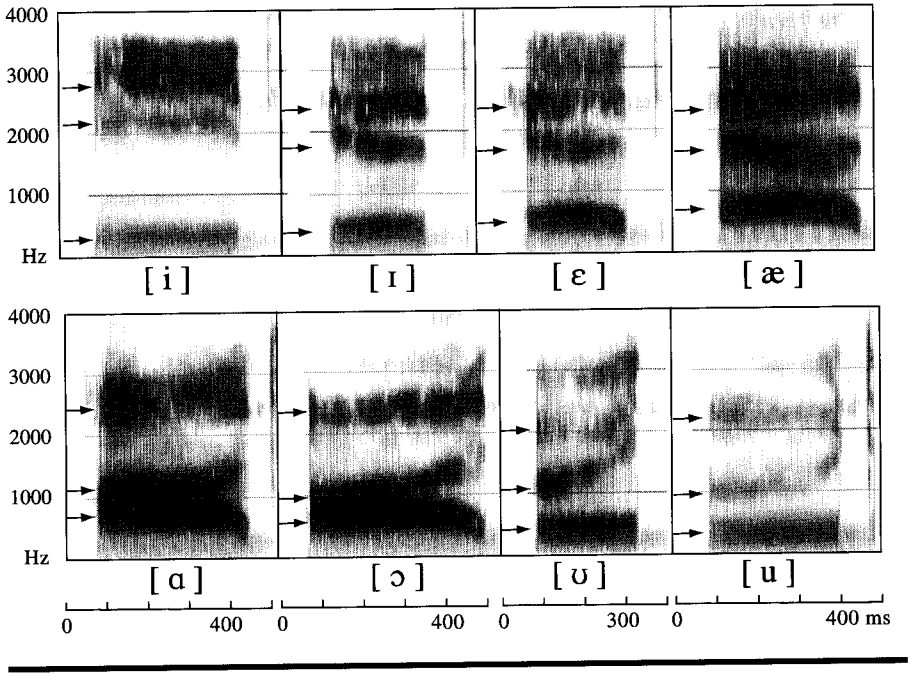


pitches in a sound. There are several free computer programs on the Web that can be used to make spectrograms. One of the best is WaveSurfer, from the Centre for Speech Technology (CTT) at KTH in Stockholm, Sweden. It is included (with permission) on the top level of the CD. You can open any of the sounds that you listen to on the CD and make spectrograms. If your computer has a built-in microphone (as on all current Macintosh computers), try recording your pronunciation of *heed, hid, head, had* and making a spectrogram.

Figure 8.3 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 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 mark, they can be seen as dark horizontal bars. The locations of the first three formants in each vowel are indicated by arrows.

FIGURE 8.3 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 8.2



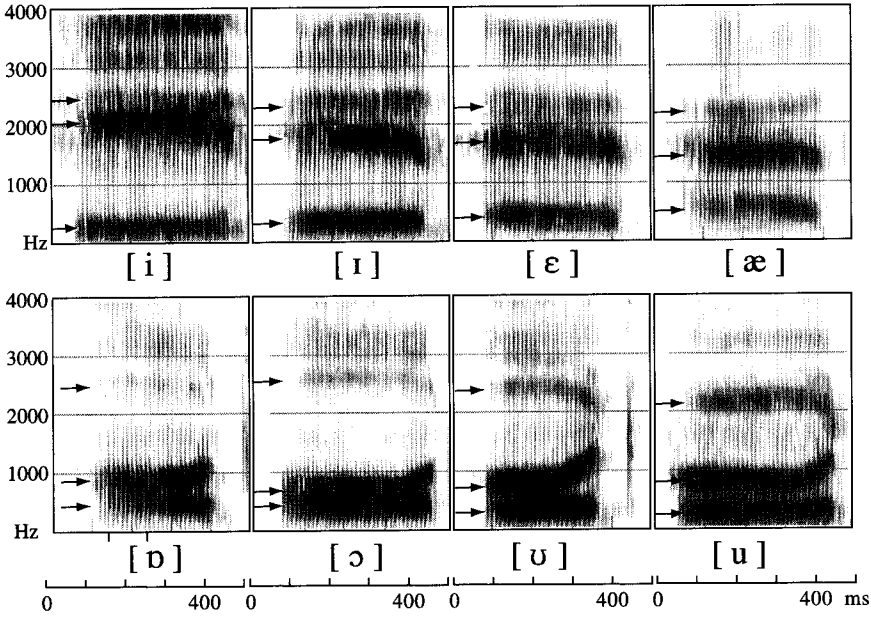
There is a great deal of similarity between Figures 8.2 and 8.3. Figure 8.2 is like a schematic spectrogram of the isolated vowels. Figure 8.3 differs in that it represents a particular American English speaker rather than the mean of a number of speakers of American English. It also shows the effects 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 at the end of [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 8.4 shows spectrograms of my form of British English. It is similar to Figure 8.3, but not exactly the same because of the differences in accent and other individual differences. My head is larger than that of the American English speaker, so all my formants are slightly lower. My vowels are less diphthongal—they have longer steady states.

CD 8.3

Whenever the vocal folds are vibrating, there are regularly spaced vertical lines, close together, on the spectrogram. During a vowel, the vertical lines are

FIGURE 8.4 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.

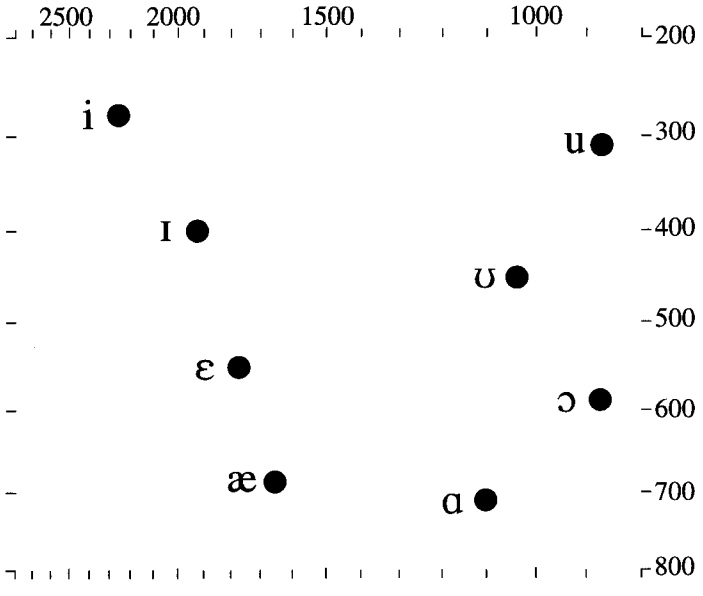


visible throughout a large part of the spectrogram. Each vertical line in the vowels is the result of the momentary increase of acoustic energy due to a single movement of the vocal folds. We have seen that it is possible to observe the pulses in a record of the waveform and from this to calculate the pitch. It is equally possible to measure the pitch from observations of the vertical striations on spectrograms. When they are close together, the pitch must be higher than when they are farther apart. At the bottom left of Figure 8.4, 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.

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 8.3 and for my vowels in Figure 8.4. In the four vowels in the bottom rows of Figures 8.3 and 8.4, 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. Lip rounding is generally characterized by the lowering of the second and third formants.

We can see some of these relationships when we plot the formant frequencies given in Figure 8.2 along axes as shown in Figure 8.5. Because the formant frequencies 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% of the energy in a vowel), the second formant scale is not as expanded as the first formant scale. (Remember that in Figures 8.3 and 8.4, 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.)

FIGURE 8.5 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.



On this kind of plot, [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, I 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 now commonly used to represent vowel qualities. To consolidate acoustic notions about vowels, you should now try to represent the vowels in Figures 8.3 and 8.4 in terms of a formant chart. I have provided arrows that mark what I 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 in Figure 8.6.

CD 8.4

FIGURE 8.6 A blank formant chart for showing the relation between vowels. Using the information in Figures 8.3 and 8.4, plot the frequency of the first formant on the ordinate (the vertical axis) and the second formant on the abscissa (the horizontal axis).