1

Articulation and Acoustics

Phonetics is concerned with describing speech. There are many different reasons for wanting to do this, which means that there are many kinds of phoneticians. Some are interested in the different sounds that occur in languages. Some study the cognitive processes involved in speaking and listening. Some are more concerned with pathological speech. Others are trying to help people speak a particular form of English. Still others are looking for ways to make computers talk more intelligibly or to get computers to recognize speech. For all these purposes, phoneticians need to find out what people are doing when they are talking and how the sounds of speech can be described.

SPEECH PRODUCTION

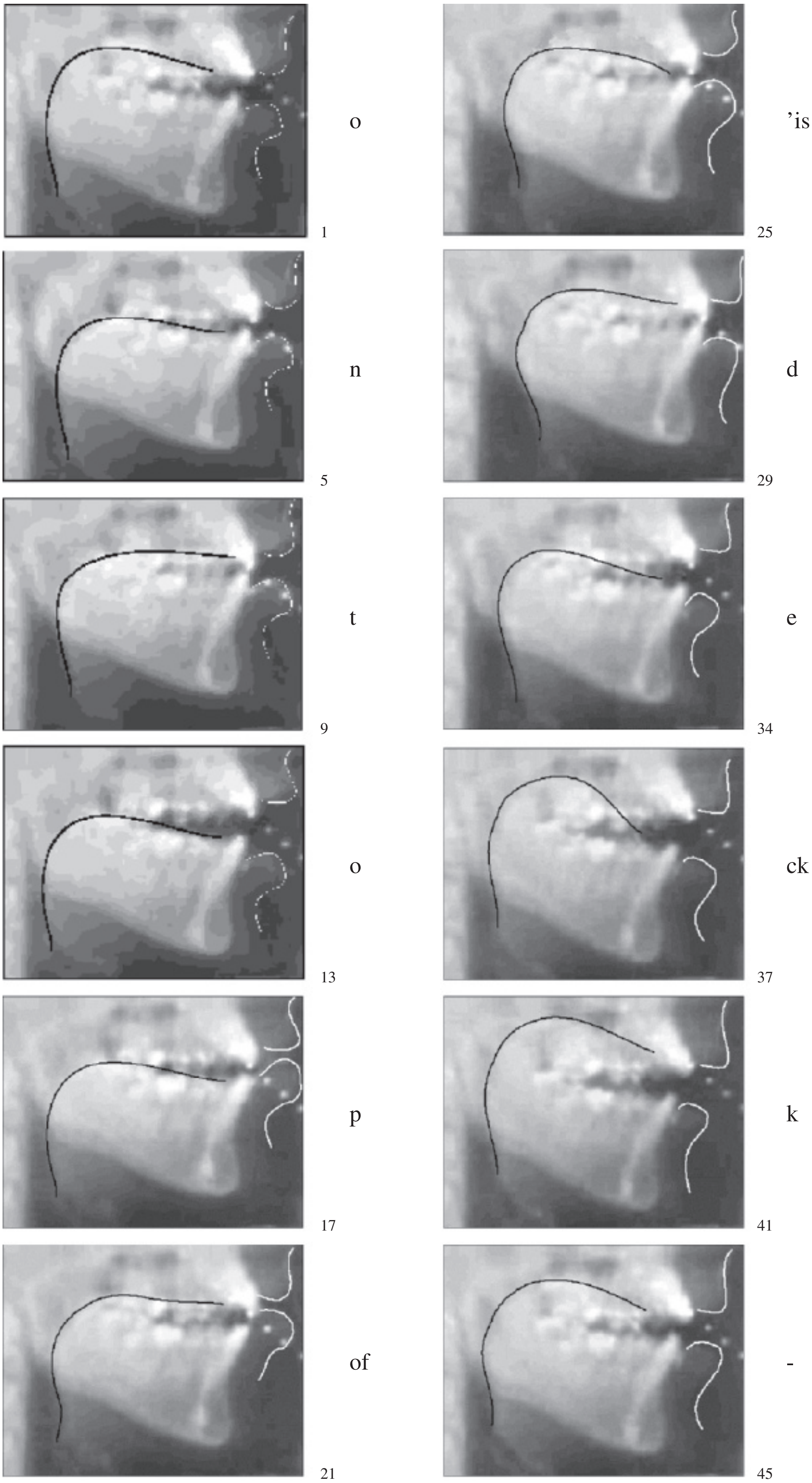
We will begin by describing how speech sounds are made. Most sounds are the result of movements of the tongue and the lips. We can think of these movements as gestures forming particular sounds. We can convey information by gestures of our hands that people can see, but in making speech that people can hear, humans have found a marvelously efficient way to impart information. The gestures of the tongue and lips are made audible so that they can be heard and recognized.

Making speech gestures audible involves pushing air out of the lungs while producing a noise in the throat or mouth. These basic noises are changed by the actions of the tongue and lips. Later, we will study how the tongue and lips make about twenty-five different gestures to form the sounds of English. We can see some of these gestures by looking at an x-ray movie (which you can watch on the book's website—http://linguistics.berkeley.edu/acip/). Figure 1.1 shows a series of frames from an x-ray movie of the phrase *on top of his deck*. In this sequence of twelve frames (one in every four frames of the movie), the tongue has been outlined to make it clearer. The lettering to the right of the frames shows, very roughly, the sounds being produced. The individual frames in the figure show that the tongue and lips move rapidly from one position to another. To appreciate how rapidly the gestures are being made, however, you should watch the movie.

EXAMPLE 1.1

Example 1.1 plays the sounds and shows the movements involved in the phrase *on top of his deck*. Even in this phrase, spoken at a normal speed, the

Figure 1.1 Frames from an x-ray movie of a speaker saying *on top of his deck*.



tongue is moving quickly. The actions of the tongue are among the fastest and most precise physical movements that people can make.

Producing any sound requires energy. In nearly all speech sounds, the basic source of power is the respiratory system pushing air out of the lungs. Try to talk while breathing in instead of out. You will find that you can do it, but it is much harder than talking when breathing out. When you talk, air from the lungs goes up the windpipe (the trachea, to use the more technical term) and into the larynx, at which point it must pass between two small muscular folds called the *vocal folds*. If the vocal folds are apart (as yours probably are right now while you are breathing in and out), the air from the lungs will have a relatively free passage into the pharynx and the mouth. But if the vocal folds are adjusted so that there is only a narrow passage between them, the airstream from the lungs will set them vibrating. Sounds produced when the vocal folds are vibrating are said to be **voiced**, as opposed to those in which the vocal folds are apart, which are said to be **voiceless**.

EXAMPLE 1.2

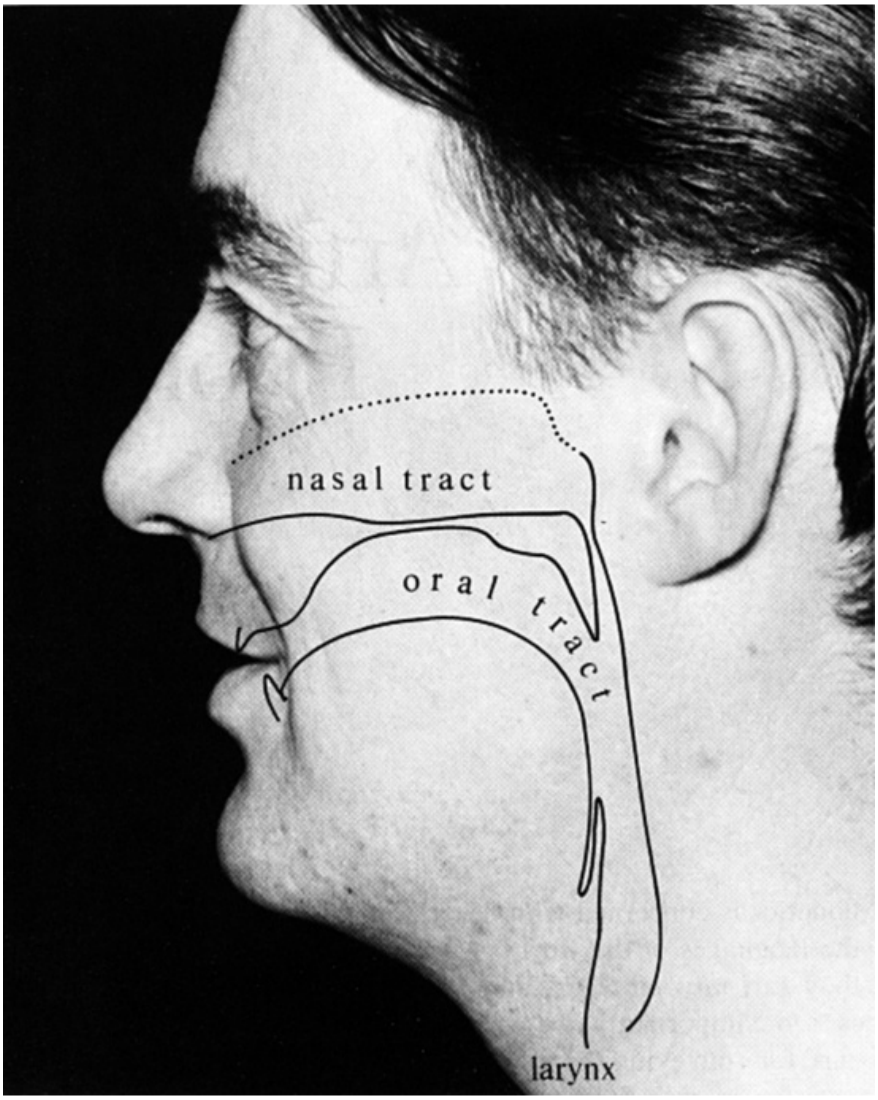
In order to hear the difference between a voiced and a voiceless sound, try saying a long '*v*' sound, which we will symbolize as [vvvvv]. Now compare this with a long '*f*' sound [fffff], saying each of them alternately—[fffffvvvvvfffffvvvvv]. (As indicated by the icon in the margin, an audio file illustrating this sequence is on the website.) Both of these sounds are formed in the same way in the mouth. The difference between them is that [v] is voiced and [f] is voiceless. You can feel the vocal fold vibrations in [v] if you put your fingertips against your larynx. You can also hear the buzzing of the vibrations in [v] more easily if you stop up your ears while contrasting [fffffvvvvv].

EXAMPLE 1.3

The difference between voiced and voiceless sounds is often important in distinguishing words. In each of the pairs of words *fat*, *vat*; *thigh*, *thy*; *Sue*, *zoo*, the first consonant in the first word of each pair is voiceless; in the second word, it is voiced. You can check this by saying just the consonant at the beginning of each of these words and try to feel and hear the voicing as suggested in the example. Try to find other pairs of words that are distinguished by one having a voiced and the other having a voiceless consonant.

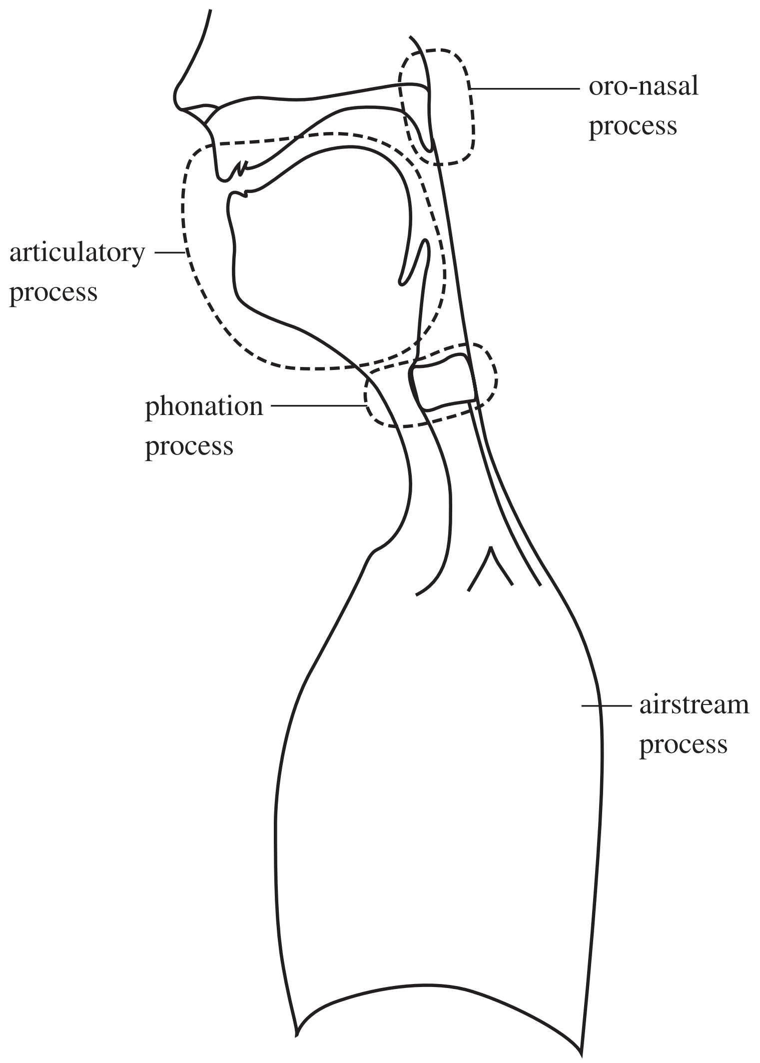
The air passages above the larynx are known as the **vocal tract**. Figure 1.2 shows their location within the head (actually, within Peter Ladefoged's head, in a photograph taken many years ago). The shape of the vocal tract is a very important factor in the production of speech, and we will often refer to a diagram of the kind that has been superimposed on the photograph in Figure 1.2. Learn to draw the vocal tract by tracing the diagram in this figure. Note that the air passages that make up the vocal tract may be divided into the oral tract, within the mouth and pharynx, and the nasal tract, within the nose. When the flap at the back of the mouth is lowered (as it probably is for you now, if you are breathing with your mouth shut), air goes in and out through the nose. Speech sounds such as [m] and [n] are produced with the vocal folds vibrating and air going out through the nose. The upper limit of the nasal tract has been marked with a dotted line since the exact boundaries of the air passages within the nose depend on soft tissues of variable size.

Figure 1.2 The vocal tract.



The parts of the vocal tract that can be used to form sounds, such as the tongue and the lips, are called *articulators*. Before we discuss them, let's summarize the speech production mechanism as a whole. Figure 1.3 (on page 6) shows the four main components—the airstream process, the phonation process, the oro-nasal process, and the articulatory process. The *airstream process* includes all the ways of pushing air out (and, as we will see later, of sucking it in) that provide the power for speech. For the moment, we have considered just the respiratory system, the lungs pushing out air, as the prime mover in this process. The *phonation process* is the name given to the actions of the vocal folds. Only two possibilities have been mentioned: voiced sounds in which the vocal folds are vibrating and voiceless sounds in which they are apart. The possibility of the airstream going out through the mouth, as in [v] or [z], or the nose, as in [m] and [n], is determined by the *oro-nasal process.* The movements of the tongue and lips interacting with the roof of the mouth and the pharynx are part of the *articulatory process*.

Figure 1.3 The four main components of the speech mechanism.



SOUND WAVES

So far, we have been describing speech sounds by stating how they are made, but it is also possible to describe them in terms of what we can hear. The way in which we hear a sound depends on its acoustic structure. We want to be able to describe the acoustics of speech for many reasons (for more on acoustic phonetics, see Keith Johnson's book *Acoustic and Auditory Phonetics*). Linguists and speech pathologists need to understand how certain sounds become confused with one another. We can give better descriptions of some sounds (such as vowels) by describing their acoustic structures rather than by describing the articulatory movements involved. Knowledge of acoustic phonetics is also helpful for understanding how computers synthesize speech and how speech recognition works (topics that are addressed more fully in Peter Ladefoged's book *Vowels and Consonants*). Furthermore, often the only permanent data that we can get of a speech event is an audio recording, as it is often impossible to obtain movies or x-rays showing what the speaker is doing. Accordingly, if we want permanent data that we can study, it will often have to come from analyzing an audio recording.

Speech sounds, like other sounds, can differ from one another in three ways. They can be the same or different in (1) pitch, (2) loudness, and (3) quality. Thus, two vowel sounds may have exactly the same pitch in the sense that they are said on the same note on the musical scale, and they may have the same loudness, yet still may differ in that one might be the vowel in *bad* and the other the vowel in *bud*. On the other hand, they might have the same vowel quality but differ in that one was said on a higher pitch or that one of them was spoken more loudly.

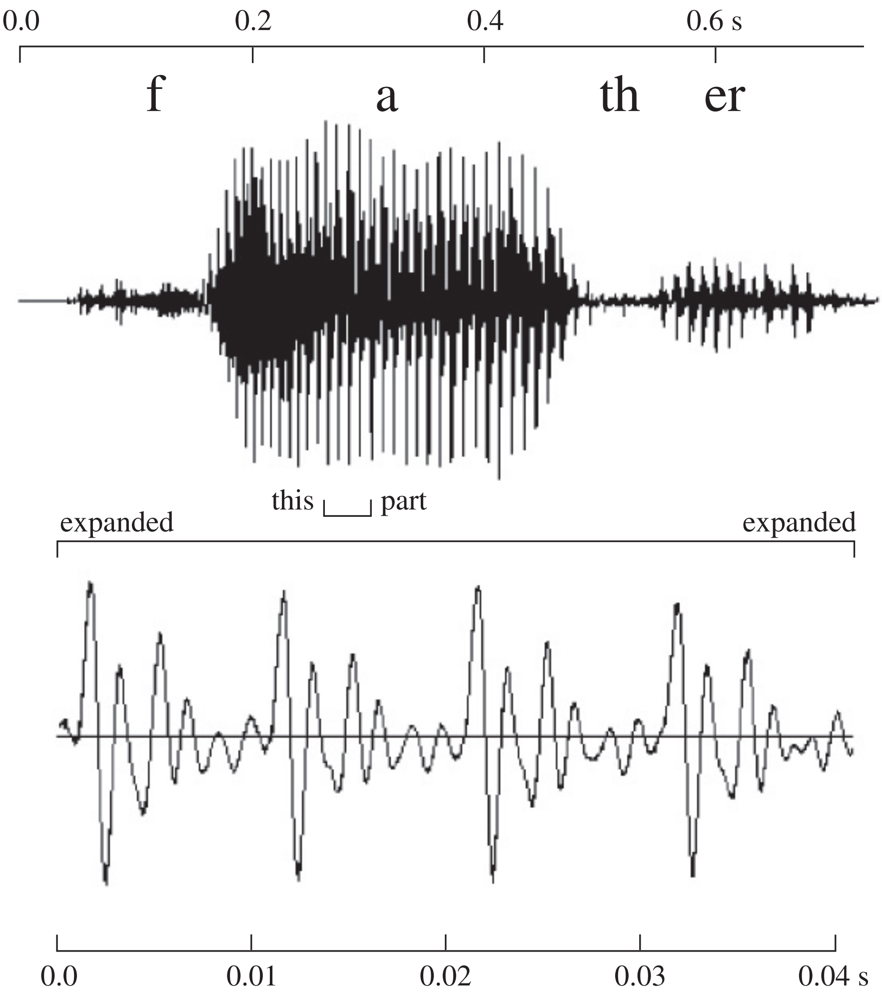
Sound consists of small variations in air pressure that occur very rapidly one after another. These variations are caused by actions of the speaker's vocal organs that are (for the most part) superimposed on the outgoing flow of lung air. Thus, in the case of voiced sounds, the vibrating vocal folds chop up the stream of lung air so that pulses of relatively high pressure alternate with moments of lower pressure. Variations in air pressure in the form of sound waves move through the air somewhat like the ripples on a pond. When they reach the ear of a listener, they cause the eardrum to vibrate. A graph of a sound wave is very similar to a graph of the movements of the eardrum.

The upper part of Figure 1.4 shows the variations in air pressure that occur during a production of the word *father*. The ordinate (the vertical axis) represents air pressure (relative to the normal surrounding air pressure), and the abscissa (the horizontal axis) represents time (relative to an arbitrary starting point). As you can see, this particular word took about 0.6 seconds to say. The lower part of the figure shows part of the first vowel in *father*. The major peaks in air pressure recur about every 0.01 seconds (i.e., every one-hundredth of a second). This is because the vocal folds were vibrating approximately one hundred times a second, producing a pulse of air every hundredth of a second. This part of the diagram shows the air pressure corresponding to four vibrations of the vocal folds. The smaller variations in air pressure that occur within each period of one-hundredth of a second are due to the way air vibrates when the vocal tract has the particular shape required for this vowel.

In the upper part of Figure 1.4, which shows the waveform for the whole word *father*, the details of the variations in air pressure are not visible because the time scale is too compressed. All that can be seen are the near-vertical lines corresponding to the individual pulses of the vocal folds. The sound [f] at the beginning of the word *father* has a low amplitude (it is not very loud, so the pressure fluctuation is not much different from zero) in comparison with the vowel that follows it, and the variations in air pressure are smaller and more nearly random. There are no regular pulses because the vocal folds are not vibrating.

The ear's response to sound is to break it down into different frequency components; in fact, fibers in the auditory nerve are tuned to specific frequencies of

Figure 1.4 The variations in air pressure that occur during Peter Ladefoged's pronunciation of the vowel in *father*.



the sound. In order to visualize what the ear hears, we simulate this property of hearing by performing a spectral analysis of the sound, which results in a **sound spectrogram**. The process is illustrated in Figure 1.5 (the web version of this figure is annotated further).

EXAMPLE 1.4

As we saw in Figure 1.4, the waveform of speech shows differences between fricative noises and vowels in both amplitude and periodicity (regular repetition of a waveform pattern). In Figure 1.5, the utterance under consideration is of the sequence [ʃɑ] as in the word "shah." The figure shows a spectral analysis of a small window of time during the "sh" sound (arrow 1 in the figure) and another during the vowel (arrow 2). The windows are shaded grey in the waveform. The spectrum shows the amplitudes of different frequency components in the sound. These amplitudes change over time, so we take spectra from many small windows of time, rotate the axes (arrows 3 and 4), and place them side to side in a three dimensional graph (the spectrogram) that shows time on the horizontal axis, frequency on the vertical axis, and amplitude in a grey scale. The spectra in

Figure 1.5 Where spectrograms come from. Small windows (shaded grey) of the acoustic waveform are spectrally analyzed, and amplitude is coded on a grey scale: darker = higher amplitude. Then the spectral slices are rotated and stacked together with each other in a three dimensional display (the spectogram). The dimensions are time on the horizontal axis, frequency on the vertical axis, and amplitude in the grey scale.

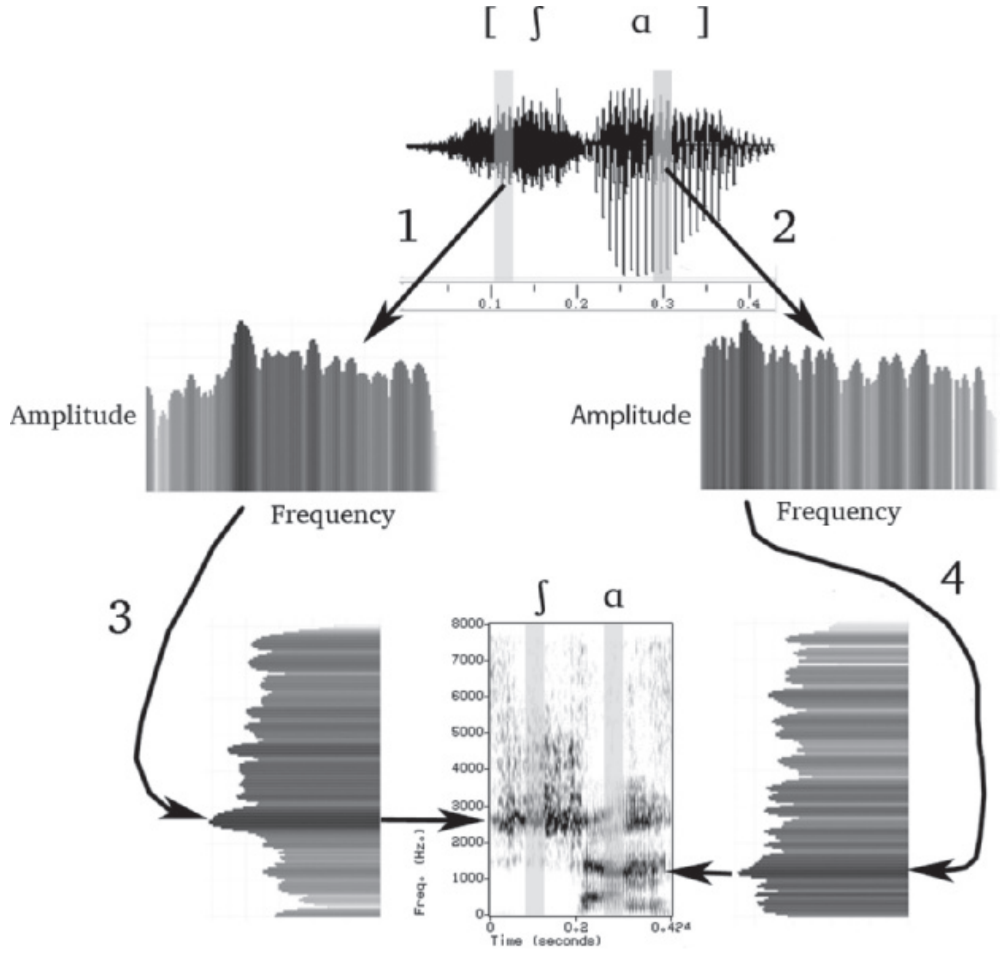


Figure 1.5 are filled with different shades of grey, depending on the height of the amplitude peak to illustrate how grey scale is used to code amplitude in spectrograms. The spectrogram in Figure 1.5 is composed of 213 spectral "slices," with one new spectrum calculated every 2 milliseconds during the utterance.

Spectrograms are a very informative visual display of speech sounds because they mimic the analysis performed by the ear. The information that we see in spectrograms is directly relevant for understanding how listeners perceive speech. Spectrograms also are very informative for deducing articulatory gestures because articulations leave acoustic "signatures" that we can see in spectrograms. For instance, in Figure 1.5, the "sh" noise is noticeably different from the noise we would expect to see in an "s" (to the trained eye). You will see quite a few spectrograms in this book, and especially on the book's website, and will become adept at reading them.

PLACES OF ARTICULATORY GESTURES

The parts of the vocal tract that can be used to form sounds are called *articulators*. The articulators that form the lower surface of the vocal tract are highly mobile. They make the gestures required for speech by moving toward the articulators that form the upper surface. Try saying the word *capital* and note the major movements of your tongue and lips. You will find that the back of the tongue moves up to make contact with the roof of the mouth for the first sound and then, comes down for the following vowel. The lips come together in the formation of *p* and then, come apart again in the vowel. The tongue tip comes up for the *t* and again, for most people, for the final *l*.

The names of the principal parts of the upper surface of the vocal tract are given in Figure 1.6. The upper lip and the upper teeth (notably the frontal incisors) are familiar enough structures. Just behind the upper teeth is a small protuberance that you can feel with the tip of the tongue. This is called the **alveolar ridge**. You can also feel that the front part of the roof of the mouth is formed by a bony structure. This is the **hard palate**. You will probably have to use a fingertip to feel farther back. Most people cannot curl the tongue up far enough to touch the **soft palate**, or **velum**, at the back of the mouth. The soft palate is a muscular flap that can be raised to press against the back wall of the pharynx and shut off the nasal tract, preventing air from going out through the nose. In this case, there is said to be a **velic closure**. This action separates the nasal tract from the oral tract so that the air can go out only through the mouth. At the lower end of the soft palate is a small appendage hanging down that is known as the *uvula*. The part of the vocal tract between the uvula and the larynx is the pharynx. The back wall of the pharynx may be considered one of the articulators on the upper surface of the vocal tract.

Figure 1.6 The principal parts of the upper surface of the vocal tract.

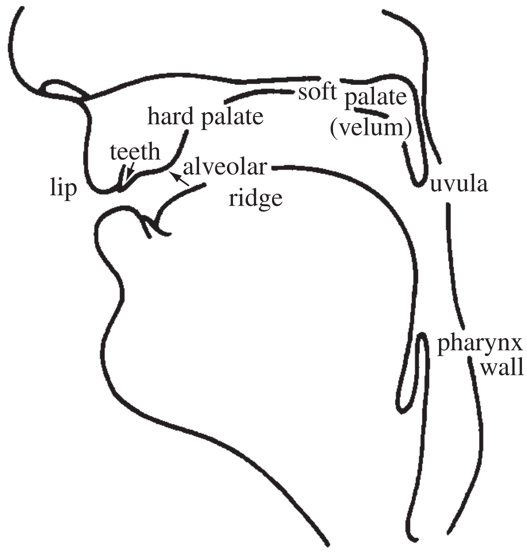


Figure 1.7 The principal parts of the lower surface of the vocal tract.

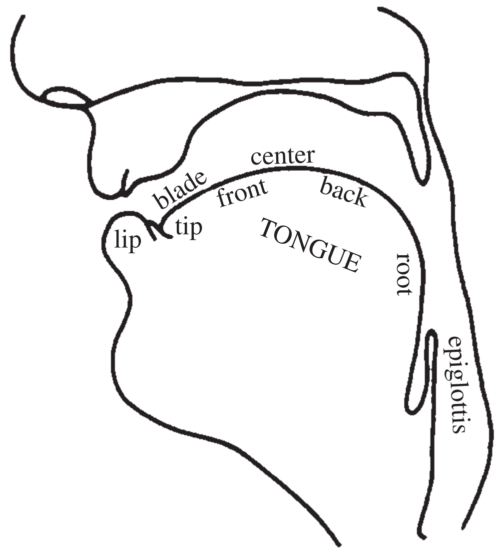


Figure 1.7 shows the lower lip and the specific names for the parts of the tongue that form the lower surface of the vocal tract. The tip and blade of the tongue are the most mobile parts. Behind the blade is what is technically called the *front* of the tongue; it is actually the forward part of the body of the tongue and lies underneath the hard palate when the tongue is at rest. The remainder of the body of the tongue may be divided into the center, which is partly beneath the hard palate and partly beneath the soft palate; the back, which is beneath the soft palate; and the root, which is opposite the back wall of the pharynx. The *epiglottis* is attached to the lower part of the root of the tongue. You may find it helpful to review the names of the articulators using the web version of Homework Exercise A for Chapter 1. It is a fun interactive way to study this information.

Bearing all these terms in mind, say the word *peculiar* and try to give a rough description of the gestures made by the vocal organs during the consonant sounds. You should find that the lips come together for the first sound. Then the back and center of the tongue are raised. But is the contact on the hard palate or on the velum? (For most people, it is centered between the two.) Then note the position in the formation of the *l*. Most people make this sound with the tip of the tongue on the alveolar ridge.

Now compare the words *true* and *tea*. In which word does the tongue movement involve a contact farther forward in the mouth? Most people make contact with the tip or blade of the tongue on the alveolar ridge when saying *tea*, but slightly farther back in *true*. Try to distinguish the differences in other consonant sounds, such as those in *sigh* and *shy* and those at the beginning of *fee* and *thief*.

When considering diagrams such as those we have been discussing, it is important to remember that they show only two dimensions. The vocal tract is a tube, and the positions of the sides of the tongue may be very different from the position of the center. In saying *sigh*, for example, there is a deep hollow in the center of the tongue that is not present when saying *shy*. We cannot represent this difference in a two-dimensional diagram that shows just the midline of the tongue—a so-called *mid-sagittal* view. We will be relying on mid-sagittal diagrams of the vocal organs to a considerable extent in this book. But we should never let this simplified view become the sole basis for our conceptualization of speech sounds.

In order to form consonants, the airstream through the vocal tract must be obstructed in some way. Consonants can be classified according to the place and manner of this obstruction. The primary articulators that can cause an obstruction in most languages are the lips, the tongue tip and blade, and the back of the tongue. Speech gestures using the lips are called **labial** articulations; those using the tip or blade of the tongue are called **coronal** articulations; and those using the back of the tongue are called **dorsal** articulations.

If we do not need to specify the place of articulation in great detail, then the articulators for the consonants of English (and of many other languages) can be described using these terms. The word *topic*, for example, begins with a coronal consonant; in the middle is a labial consonant; and at the end is a dorsal consonant. Check this by feeling that the tip or blade of your tongue is raised for the first (coronal) consonant, your lips close for the second (labial) consonant, and the back of your tongue is raised for the final (dorsal) consonant.

These terms, however, do not specify articulatory gestures in sufficient detail for many phonetic purposes. We need to know more than which articulator is making the gesture, which is what the terms *labial*, *coronal*, and *dorsal* tell us. We also need to know what part of the upper vocal tract is involved. More specific places of articulation are indicated by the arrows going from one of the lower articulators to one of the upper articulators in Figure 1.8. Because there are so many possibilities in the coronal region, this area is shown in more detail at the right of the figure. The principal terms for the particular types of obstruction required in the description of English are as follows.

1. **Bilabial**

   (Made with the two lips.) Say words such as *pie, buy, my* and note how the lips come together for the first sound in each of these words. Find a comparable set of words with bilabial sounds at the end.

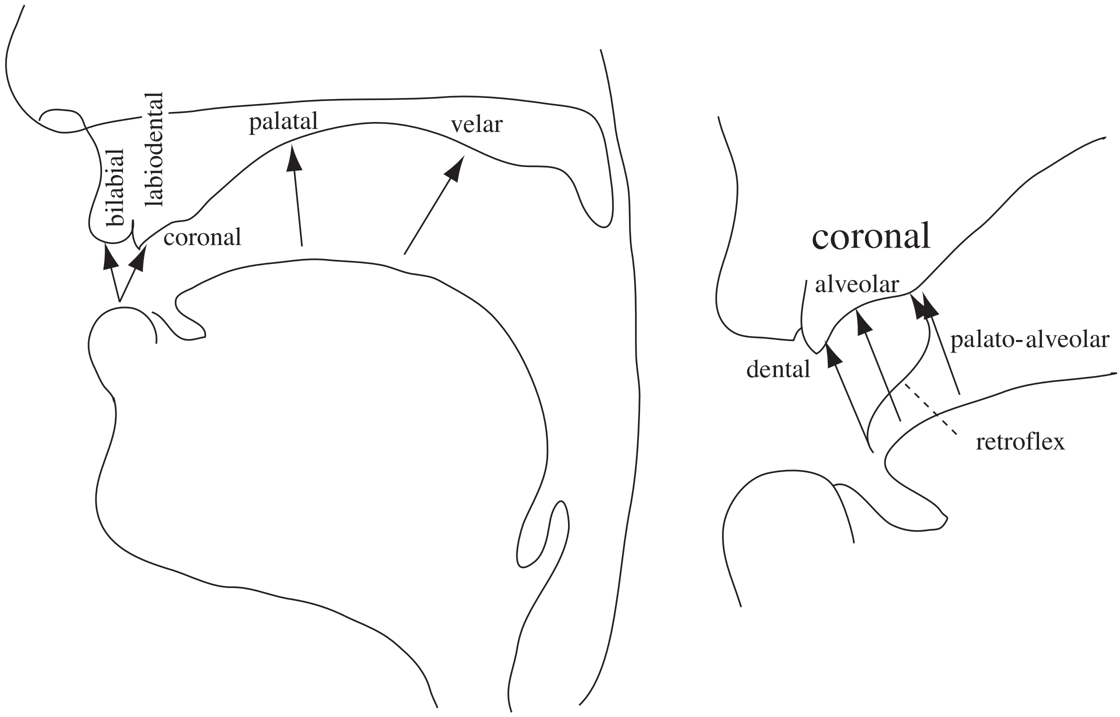
2. **Labiodental**

   (Lower lip and upper front teeth.) Most people, when saying words such as *fie* and *vie*, raise the lower lip until it nearly touches the upper front teeth.

3. **Dental**

   (Tongue tip or blade and upper front teeth.) Say the words *thigh* and *thy*. Some people (most speakers of American English as spoken in the Midwest and on the West Coast) have the tip of the tongue protruding between the upper and lower front teeth; others (most speakers of British English)

Figure 1.8 A sagittal section of the vocal tract, showing the places of articulation that occur in English. The coronal region is shown in more detail at the right.



have it close behind the upper front teeth. Both sounds are normal in English, and both may be called *dental*. If a distinction is needed, sounds in which the tongue protrudes between the teeth may be called **interdental**.

4. **Alveolar**

   (Tongue tip or blade and the alveolar ridge.) Again there are two possibilities in English, and you should find out which you use. You may pronounce words such as *tie*, *die*, *nigh*, *sigh*, *zeal*, *lie* using the tip of the tongue or the blade of the tongue. You may use the tip of the tongue for some of these words and the blade for others. For example, some people pronounce [s] with the tongue tip tucked behind the lower teeth, producing the constriction at the alveolar ridge with the blade of the tongue; others have the tongue tip up for [s]. Feel how you normally make the alveolar consonants in each of these words, and then try to make them in the other way. A good way to appreciate the difference between dental and alveolar sounds is to say *ten* and *tenth* (or *n* and *nth*). Which *n* is farther back? (Most people make the one in *ten* on the alveolar ridge and the one in *tenth* as a dental sound with the tongue touching the upper front teeth.)

5. **Retroflex**

   (Tongue tip and the back of the alveolar ridge.) Many speakers of English do not use retroflex sounds at all. But some speakers begin words such as *rye*, *row*, *ray* with retroflex sounds. Note the position of the tip of your tongue in these words. Speakers who pronounce *r* at the ends of words may also have retroflex sounds with the tip of the tongue raised in words such as *ire*, *hour*, and *air*.

6. **Post-Alveolar**

   (Tongue blade and the back of the alveolar ridge.) Say words such as *shy*, *she*, and *show*. During the consonants, the tip of your tongue may be down behind the lower front teeth or up near the alveolar ridge, but the blade of the tongue is always close to the back part of the alveolar ridge. Because these sounds are made at the boundary between the alveolar ridge and the hard palate, they can also be called **palato-alveolar**. It is possible to pronounce them with either the tip or blade of the tongue. Try saying *shipshape* with your tongue tip up on one occasion and down on another. Note that the blade of the tongue will always be raised. You may be able to feel the place of articulation more distinctly if you hold the position while taking in a breath through the mouth. The incoming air cools the region where there is greatest narrowing, the blade of the tongue and the back part of the alveolar ridge.

7. **Palatal**

   (Front of the tongue and hard palate.) Say the word *you* very slowly so that you can isolate the consonant at the beginning. If you say this consonant by itself, you should be able to feel that it begins with the front of the tongue raised toward the hard palate. Try to hold the beginning consonant position and breathe in through the mouth. You will probably be able to feel the rush of cold air between the front of the tongue and the hard palate.

8. **Velar**

   (Back of the tongue and soft palate.) The consonants that have the place of articulation farthest back in English are those that occur at the end of words such as *hack*, *hag*, and *hang*. In all these sounds, the back of the tongue is raised so that it touches the velum.

As you can tell from the descriptions of these articulatory gestures, the first two, bilabial and labiodental, can be classified as labial, involving at least the lower lip; the next four—dental, alveolar, retroflex, and palato-alveolar (post-alveolar)—are coronal articulations, with the tip or blade of the tongue raised; and the last, velar, is a dorsal articulation, using the back of the tongue. Palatal sounds are sometimes classified as coronal articulations and sometimes as dorsal articulations, a point to which we shall return.

To get the feeling of different places of articulation, consider the consonant at the beginning of each of the following words: *fee*, *theme*, *see*, *she.* Say these consonants by themselves. Are they voiced or voiceless? Now note that the place of articulation moves back in the mouth in making this series of voiceless consonants, going from labiodental, through dental and alveolar, to palato-alveolar.

THE ORO-NASAL PROCESS

Consider the consonants at the ends of *rang*, *ran*, and *ram.* When you say these consonants by themselves, note that the air is coming out through the nose. In the formation of these sounds in sequence, the point of articulatory closure moves forward, from velar in *rang*, through alveolar in *ran*, to bilabial in *ram.* In each case, the air is prevented from going out through the mouth but is able to go out through the nose because the soft palate, or velum, is lowered.

In most speech, the soft palate is raised so that there is a velic closure. When it is lowered and there is an obstruction in the mouth, we say that there is a nasal consonant. Raising or lowering the velum controls the oro-nasal process, the distinguishing factor between oral and nasal sounds.

MANNERS OF ARTICULATION

At most places of articulation, there are several basic ways in which articulatory gestures can be accomplished. The articulators may close off the oral tract for an instant or a relatively long period; they may narrow the space considerably; or they may simply modify the shape of the tract by approaching each other.

Stop

(Complete closure of the articulators involved so that the airstream cannot escape through the mouth.) There are two possible types of stop.

Oral stop If, in addition to the articulatory closure in the mouth, the soft palate is raised so that the nasal tract is blocked off, then the airstream will be completely obstructed. Pressure in the mouth will build up and an **oral stop** will be formed. When the articulators come apart, the airstream will be released in a small burst of sound. This kind of sound occurs in the consonants in the words *pie*, *buy* (bilabial closure), *tie*, *dye* (alveolar closure), and *kye*, *guy* (velar closure). Figure 1.9 shows the positions of the vocal organs in the bilabial stop in *buy*. These sounds are called **plosives** in the International Phonetic Association's (IPA's) alphabet (see inside the front cover of this book).

Nasal stop If the air is stopped in the oral cavity but the soft palate is down so that air can go out through the nose, the sound produced is a **nasal stop**. Sounds of this kind occur at the beginning of the words *my* (bilabial closure) and *nigh* (alveolar closure), and at the end of the word *sang* (velar closure). Figure 1.10 (on page 16) shows the position of the vocal organs during the bilabial nasal stop in *my.* Apart from the presence of a velic opening, there is no difference between this stop and the one in *buy* shown in Figure 1.9 (on page 16). Although both the nasal sounds and the oral sounds can be classified as stops, the term **stop** by itself is almost always used by phoneticians to indicate an oral stop, and the term **nasal** to indicate a nasal stop. Thus, the consonants at the beginnings of the words *day*

Figure 1.9 The positions of the vocal organs in the bilabial stop in *buy*.

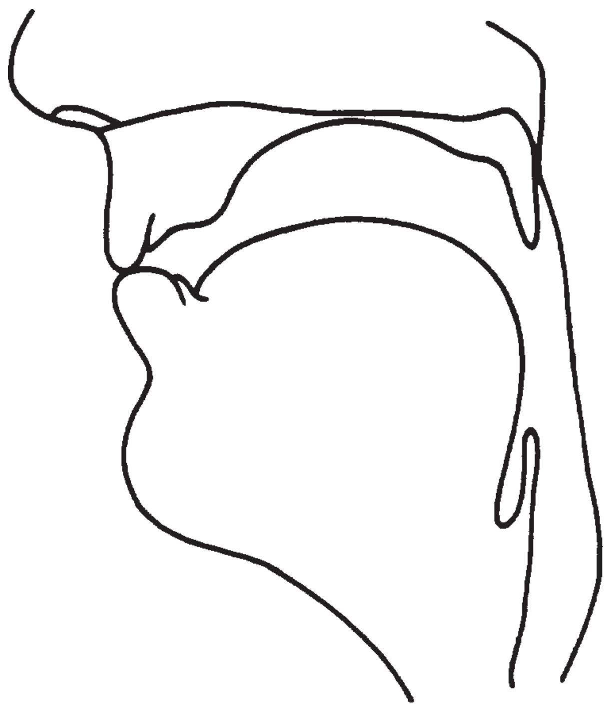
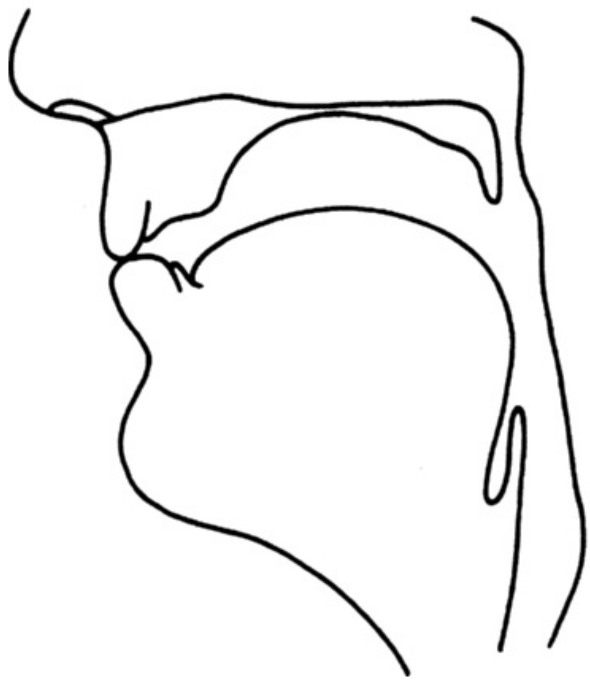
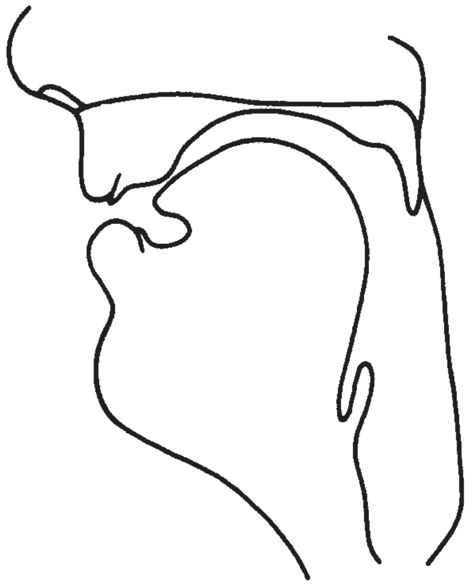


Figure 1.10 The positions of the vocal organs in the bilabial nasal (stop) in *my*.



and *neigh* would be called an alveolar stop and an alveolar nasal, respectively. Although the term **stop** may be defined so that it applies only to the prevention of air escaping through the mouth, it is commonly used to imply a complete stoppage of the airflow through both the nose and the mouth.

Figure 1.11 The positions of the vocal organs in the palato-alveolar (post-alveolar) fricative in *shy*.



Fricative

(Close approximation of two articulators so that the airstream is partially obstructed and turbulent airflow is produced.) The mechanism involved in making these slightly hissing sounds may be likened to that involved when the wind whistles around a corner. The consonants in *fie*, *vie* (labiodental), *thigh*, *thy* (dental), *sigh*, *zoo* (alveolar), and *shy* (palato-alveolar) are examples of fricative sounds. Figure 1.11 illustrates one pronunciation of the palato-alveolar fricative consonant in *shy.* Note the narrowing of the vocal tract between the blade of the tongue and the back part of the alveolar ridge. The higher-pitched sounds with a more obvious hiss, such as those in *sigh* and *shy*, are sometimes called **sibilants**.

Approximant

(A gesture in which one articulator is close to another, but without the vocal tract being narrowed to such an extent that a turbulent airstream is produced.) In saying the first sound in *yacht*, the front of the tongue is raised toward the palatal area of the roof of the mouth, but it does not come close enough for a fricative sound to be produced. The consonants in the word *we* (approximation between the lips and in the velar region) and, for some people, in the word *raw* (approximation in the alveolar region) are also examples of approximants.

Lateral (Approximant)

(Obstruction of the airstream at a point along the center of the oral tract, with incomplete closure between one or both sides of the tongue and the roof of the mouth.) Say the word *lie* and note how the tongue touches near the center of the alveolar ridge. Prolong the initial consonant and note how, despite the closure formed by the tongue, air flows out freely, over the side of the tongue. Because there is no stoppage of the air, and not even any fricative noises, these sounds are classified as approximants. The consonants in words such as *lie* and *laugh* are alveolar lateral approximants, but they are usually called just *alveolar laterals*, their approximant status being assumed. You may be able to find out which side of the tongue is not in contact with the roof of the mouth by holding the consonant position while you breathe inward. The tongue will feel colder on the side that is not in contact with the roof of the mouth.

Additional Consonantal Gestures

In this preliminary chapter, it is not necessary to discuss all of the manners of articulation used in the various languages of the world—nor, for that matter, in English. But it might be useful to know the terms **trill** (sometimes called **roll**) and **tap** (sometimes called **flap**). Tongue-tip trills occur in some forms of Scottish English in words such as *rye* and *raw*, and as the "rolled *r*" in Spanish. Taps, in which the tongue makes a single tap against the alveolar ridge, occur in the middle of a word such as *pity* in many forms of American English.

The production of some sounds involves more than one of these manners of articulation. Say the word *cheap* and think about how you make the first sound. At the beginning, the tongue comes up to make contact with the back part of the alveolar ridge to form a stop closure. This contact is then slackened so that there is a fricative at the same place of articulation. This kind of combination of a stop immediately followed by a fricative is called an **affricate**, in this case a palato-alveolar (or post-alveolar) affricate. There is a voiceless affricate at the beginning and end of the word *church*. The corresponding voiced affricate occurs at the beginning and end of *judge*. In all these sounds the articulators (tongue tip or blade and alveolar ridge) come together for the stop and then, instead of coming fully apart, separate only slightly so that a fricative is made at approximately the same place of articulation. Try to feel these movements in your own pronunciation of these words.

Words in English that start with a vowel in the spelling (like *eek*, *oak*, *ark*, etc.) are pronounced with a **glottal stop** at the beginning of the vowel. This "glottal catch" sound isn't written in these words and is easy to overlook; but in a sequence of two words in which the first word ends with a vowel and the second starts with a vowel, the glottal stop is sometimes obvious. For example, the phrase *flee east* is different from the word *fleeced* in that the first has a glottal stop at the beginning of *east*.

To summarize, the consonants we have been discussing so far may be described in terms of five factors:

1. state of the vocal folds (voiced or voiceless);
2. place of articulation;

3. central or lateral articulation;
4. soft palate raised to form a velic closure (oral sounds) or lowered (nasal sounds); and
5. manner of articulatory action.

Thus, the consonant at the beginning of the word *sing* is a (1) voiceless, (2) alveolar, (3) central, (4) oral, (5) fricative; and the consonant at the end of *sing* is a (1) voiced, (2) velar, (3) central, (4) nasal, (5) stop.

On most occasions, it is not necessary to state all five points. Unless a specific statement to the contrary is made, consonants are usually presumed to be central, not lateral, and oral rather than nasal. Consequently, points (3) and (4) may often be left out, so the consonant at the beginning of *sing* is simply called a voiceless alveolar fricative. When describing nasals, point (4) has to be specifically mentioned and point (5) can be left out, so the consonant at the end of *sing* is simply called a voiced velar nasal.

THE ACOUSTICS OF CONSONANTS

At this stage, we will not go too deeply into the acoustics of consonants, simply noting a few distinctive points about them from waveforms and spectrograms. The places of articulation are not obvious in any waveform, but the differences in some of the principal manners of articulation—stop, nasal, fricative, and approximant—are usually apparent. Furthermore, as already pointed out, you can also see the differences between voiced and voiceless sounds. Place of articulation is more apparent in spectrograms, though it can still be a surprisingly subtle feature given how important it is for speech communication.

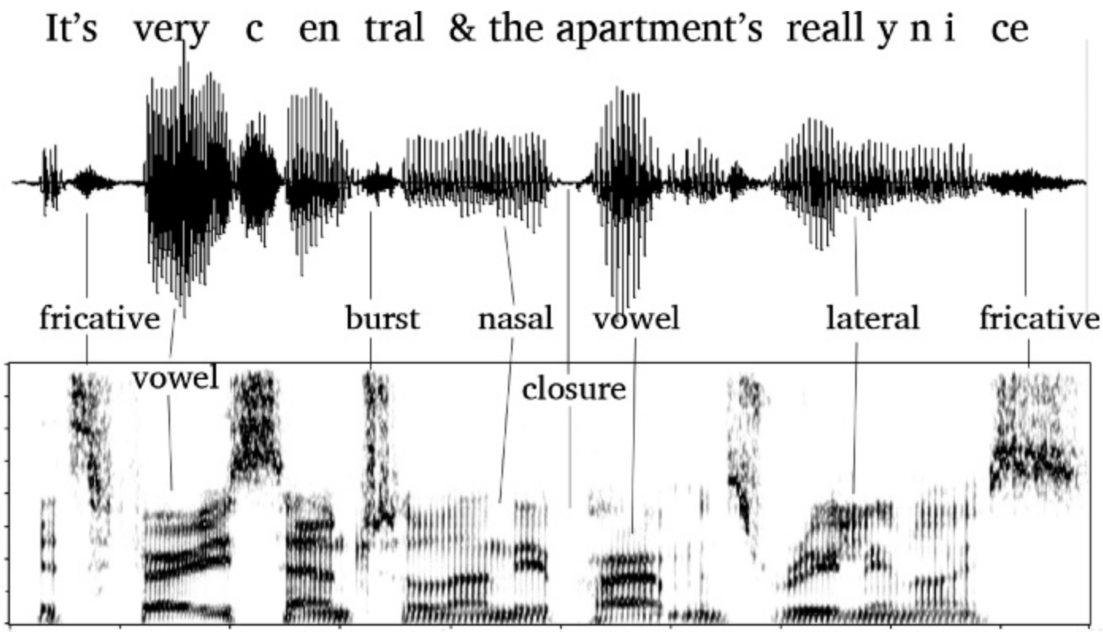
The top part of Figure 1.12 (on page 20) shows the waveform of the phrase *It's very central and the apartment is really nice*, labeled roughly in ordinary spelling. The lower part shows a spectrogram of the same utterance (this utterance is one of several contributed by a Dubliner—listen to it in the "Extras" section of the website for this book). This phrase took about two seconds.

WEB Extras

Looking at the labels in the middle of the figure, you can see that fricatives (in *it's* and *nice*) are very different from vowels in both the waveform and in the spectrogram. But notice that the nasal in *and* and the lateral in *really* are much easier to see in the spectrogram than in the waveform.

Try to guess where the "n" of *nice* is in the waveform and then look for it in the spectrogram. Even though you don't know exactly what to look for in the spectrogram, it is probably apparent that something changed in the spectrogram (there is less energy, hence less ink, as if someone took an eraser and made a vertical white stripe), where it would be next to impossible to identify the nasal in the waveform.

Figure 1.12 The waveform and spectrogram of the phrase *It's very central and the apartment's really nice.*

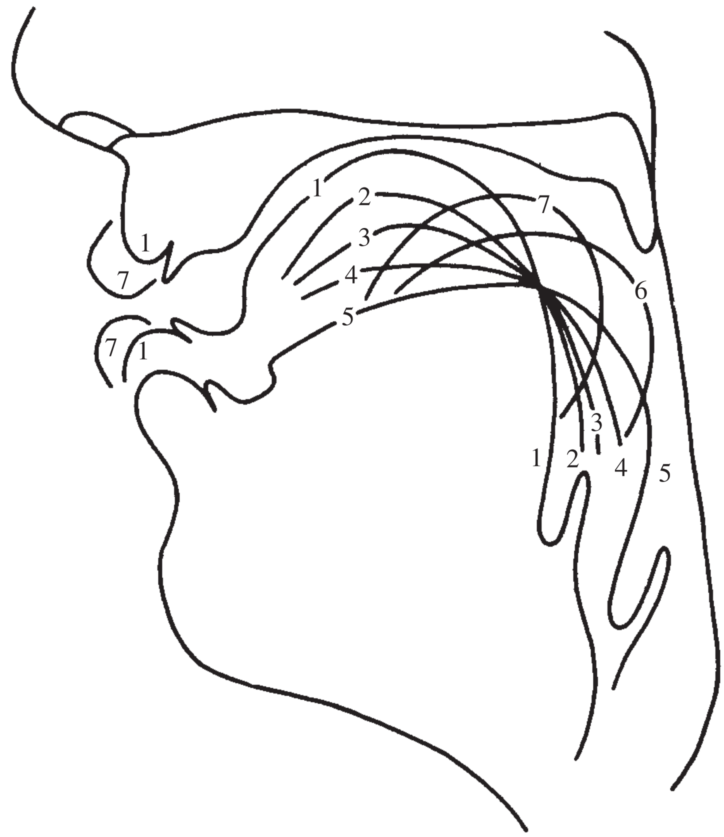


THE ARTICULATION OF VOWEL SOUNDS

In the production of vowel sounds, the articulators do not come very close together, and the passage of the airstream is relatively unobstructed. For this reason, it is much more difficult to feel the position of the tongue during vowel sounds, than in consonants. We can describe vowel sounds roughly in terms of the position of the highest point of the tongue and the position of the lips. (As we will see later, more accurate descriptions can be made in acoustic terms.) Figure 1.13 shows the articulatory position for the vowels in *heed*, *hid*, *head*, *had*, *father*, *good*, *food.* Of course, in saying these words, the tongue and lips are in continuous motion throughout the vowels, as we saw in the x-ray movie in example 1.1 on the book's website. The positions shown in the figure are best considered as the targets of the gestures for the vowels.

As you can see, in all these vowel gestures, the tongue tip is down behind the lower front teeth, and the body of the tongue is domed upward. Check that this is so in your own pronunciation. You will notice that you can prolong the [h] sound and that there is no mouth movement between the [h] and the following vowel; the [h] is like a voiceless version of the vowel that comes after it. In the first four vowels, the highest point of the tongue is in the front of the mouth. Accordingly, these vowels are called **front vowels**. The tongue is fairly close to the roof of the mouth for the vowel in *heed* (you can feel that this is so by breathing inward while holding the target position for this vowel), slightly less close for the vowel in *hid* (for this and most other vowels it is difficult to localize the position by breathing inward; the articulators are too far apart), and

Figure 1.13 The positions of the vocal organs for the vowels in the words 1 *heed*, 2 *hid*, 3 *head*, 4 *had*, 5 *father*, 6 *good*, 7 *food*. The lip positions for vowels 2, 3, and 4 are between those shown for 1 and 5. The lip position for vowel 6 is between those shown for 1 and 7.



lower still for the vowels in *head* and *had.* If you look in a mirror while saying the vowels in these four words, you will find that the mouth becomes progressively more open while the tongue remains in the front of the mouth. The vowel in *heed* is classified as a *high front vowel*, and the vowel in *had* as a *low front vowel*. The height of the tongue for the vowels in the other words is between these two extremes, and they are therefore called *mid-front vowels*. The vowel in *hid* is a mid-high vowel, and the vowel in *head* is a mid-low vowel.

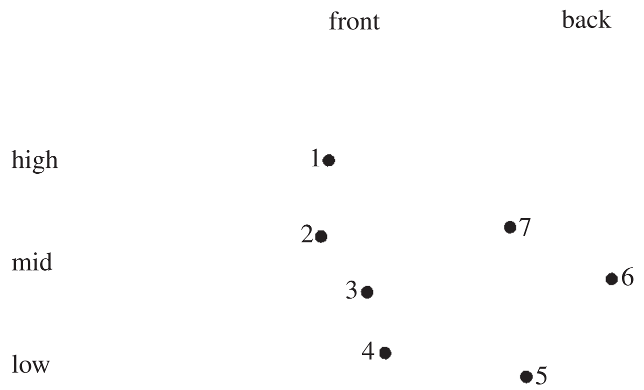
Now try saying the vowels in *father*, *good*, *food.* Figure 1.13 also shows the articulatory targets for these vowels. In all three, the tongue is close to the back surface of the vocal tract. These vowels are classified as **back vowels**. The body of the tongue is highest in the vowel in *food* (which is therefore called a *high back vowel*) and lowest in the first vowel in *father* (which is therefore called a *low back vowel*). The vowel in *good* is a mid-high back vowel. The tongue may be near enough to the roof of the mouth for you to be able to feel the rush of cold air when you breathe inward while holding the position for the vowel in *food.*

Lip gestures vary considerably in different vowels. They are generally closer together in the mid-high and high back vowels (as in *good*, *food*), though in some forms of American English this is not so. Look at the position of your lips in a mirror while you say just the vowels in *heed*, *hid*, *head*, *had*, *father*, *good*, *food.* You will probably find that in the last two words, there is a movement of the lips in addition to the movement that occurs because of the lowering and raising of the jaw. This movement is called *lip rounding*. It is usually most noticeable in the inward movement of the corners of the lips. Vowels may be described as being **rounded** (as in *who'd*) or **unrounded** (as in *heed*).

In summary, the targets for vowel gestures can be described in terms of three factors: (1) the height of the body of the tongue; (2) the front–back position of the tongue; and (3) the degree of lip rounding. The relative positions of the highest points of the tongue are given in Figure 1.14. Say just the vowels in the words given in the figure caption and check that your tongue moves in the pattern described by the points. It is very difficult to become aware of the position of the tongue in vowels, but you can probably get some impression of tongue height by observing the position of your jaw while saying just the vowels in the four words *heed*, *hid*, *head*, *had.* You should also be able to feel the difference between front and back vowels by contrasting words such as *he* and *who.* Say these words silently and concentrate on the sensations involved. You should feel the tongue going from front to back as you say *he*, *who.* You can also feel your lips becoming more rounded.

As you can see from Figure 1.14, the specification of vowels in terms of the position of the highest point of the tongue is not entirely satisfactory for a number of reasons. First, the vowels classified as high do not have the same tongue height. The back high vowel (point 7) is nowhere near as high as the front vowel (point 1). Second, the so-called back vowels vary considerably in their degree of "backness." Third, as you can see by looking at Figure 1.13, this kind of specification disregards considerable differences in the shape of the tongue in front vowels and in back vowels. Nor does it take into account the width of the pharynx, which varies considerably and is not entirely dependent on the height of the tongue in different vowels. We will discuss better ways of describing vowels in Chapters 4 and 9.

Figure 1.14 The relative positions of the highest points of the tongue in the vowels in 1 *heed*, 2 *hid*, 3 *head*, 4 *had*, 5 *father*, 6 *good*, 7 *food*.



THE SOUNDS OF VOWELS

Studying the sounds of vowels requires a greater knowledge of acoustics than we can handle at this stage of the book. We can, however, note some comparatively straightforward facts about vowel sounds. Vowels, like all sounds except the pure tone of a tuning fork, have complex structures. We can think of them as containing a number of different pitches simultaneously. There is the pitch at which the vowel is actually spoken, which depends on the pulses being produced by the vibrating vocal folds; and, quite separate from this, there are overtone pitches that depend on the shape of the resonating cavities of the vocal tract. These overtone pitches give the vowel its distinctive quality. Look back at Figure 1.12 to see what these overtone pitches look like in a spectrogram. The second vowel in *apartment* is labeled in that figure, and when you look at it in the spectrogram one key visual feature is that there are three dark bands running horizontally through the vowel. These bands occur at different frequencies for different vowels. We will enlarge on this notion in Chapter 8. Here we will consider briefly how one vowel is distinguished from another by the pitches of the overtones.

Normally, one cannot hear the separate overtones of a vowel as distinguishable pitches. The only sensation of pitch is the note on which the vowel is said, which depends on the rate of vibration of the vocal folds. But there are circumstances in which the overtones of each vowel can be heard. Try saying just the vowels in the words *heed*, *hid*, *head*, *had*, *hod*, *hawed*, *hood*, *who'd*, making all of them long vowels. Now whisper these vowels. When you whisper, the vocal folds are not vibrating, and there is no regular pitch of the voice. Nevertheless, you can hear that this set of vowels forms a series of sounds on a continuously descending pitch. What you are hearing corresponds to a group of overtones that characterize the vowels. These overtones are highest for the vowel in *heed* and lowest for the vowel in either *hawed*, *hood*, or *who'd*. Which of the three vowels is the lowest depends on your regional accent. Accents of English differ slightly in the pronunciation of these vowels. In the audio examples that accompany this chapter, you can hear these vowels whispered.

There is another way to produce something similar to this whispered pitch. Try whistling a very high note, and then the lowest note that you can. You will find that for the high note you have to have your tongue in the position for the vowel in *heed*, and for the low note your tongue is in the position for one of the vowels in *hawed*, *hood*, *who'd*. From this, it seems as if there is some kind of high pitch associated with the high front vowel in *heed,* and a low pitch associated with one of the back vowels. The lowest whistled note corresponds to the tongue and lip gestures very much like those used for the vowel in *who*. A good way to learn how to make a high back vowel is to whistle your lowest note possible, and then add voicing. EXAMPLE 1.5

Another way of minimizing the sound of the vocal fold vibrations is to say the vowels in a very low, creaky voice (listen to example 1.4 in the supplemental materials). It is easiest to produce this kind of voice with a vowel such as that in *had* or *hod.* Some people can produce a creaky-voice sound in which the rate of vibration of the vocal folds is so low that you can hear the individual pulsations.

Try saying just the vowels in *had*, *head*, *hid*, *heed* in a creaky voice. You should be able to hear a change in pitch, although, in one sense, the pitch of all of them is just that of the low, creaky voice. When saying the vowels in the order *heed*, *hid*, *head*, *had,* you can hear a sound that steadily increases in pitch by approximately equal steps with each vowel. Now say the vowels in *hod*, *hood*, *who'd* in a creaky voice. These three vowels have overtones with a steadily decreasing pitch. Example 1.4 has an audio file of Peter Ladefoged saying the vowels in the words *heed*, *hid*, *head*, *had*, *hod*, *hawed*, *hood*, *who'd* in his British accent. The first four of these vowels have a quality that clearly goes up in pitch, and the last four have a declining pitch.

In summary, vowel sounds may be said on a variety of notes (voice pitches), but they are distinguished from one another by two characteristic vocal tract pitches associated with their overtones. One of them (actually the higher of the two) goes downward throughout most of the series *heed*, *hid*, *head*, *had*, *hod*, *hawed*, *hood*, *who'd* and corresponds roughly to the difference between front and back vowels. The other is low for vowels in which the tongue position is high and high for vowels in which the tongue position is low. It corresponds (inversely) to what we called *vowel height* in articulatory terms. These characteristic overtones are called the **formants** of the vowels. The one with the lower pitch (distinguishable in creaky voice) is called the *first formant*, and the higher one (the one heard when whispering) is the *second formant*.

The notion of a formant (actually the second formant) distinguishing vowels has been known for a long time. It was observed by Isaac Newton, who, in about 1665, wrote in his notebook: "The filling of a very deepe flaggon with a constant streame of beere or water sounds ye vowells in this order *w*, *u*, *ω*, *o*, *a*, *e*, *i*, *y*." He was about twelve years old at the time. (The symbols used here are the best matches to the letters in Newton's handwriting in his notebook, which is in the British Museum. They probably refer to the vowels in words such as *woo*, *hoot*, *foot*, *coat*, *cot*, *bait*, *bee*, *ye*.) Fill a deep narrow glass with water (or beer!) and see if you can hear something like the second formant in the vowels in these words as the glass fills up.

SUPRASEGMENTALS

Vowels and consonants can be thought of as the segments of which speech is composed. Together they form the syllables that make up utterances. Superimposed on the syllables are other features known as *suprasegmentals*. These include variations in stress and pitch. Variations in length are also usually considered to be suprasegmental features, although they can affect single segments as well as whole syllables. We will defer detailed descriptions of the articulation and the corresponding acoustics of these aspects of speech till later in this book.

Variations in stress are used in English to distinguish between a noun and a verb, as in *(an) insult* versus *(to) insult.* Say these words yourself, and check which syllable has the greater stress. Then compare similar pairs such as *(a) pervert*, *(to) pervert* or *(an) overflow*, *(to) overflow.* Listen to these words in the supplemental materials. You should find that in the nouns, the stress is on the first syllable, but in the verbs, it is on the last. Thus, stress can have a grammatical function in English. It can also be used for contrastive emphasis (as in *I want a **red** pen, not a **black** one*). Stress in English is produced by (1) increased activity of the respiratory muscles, producing greater loudness, as well as by (2) exaggeration of consonant and vowel properties such as vowel height and stop aspiration, and (3) exaggeration of pitch so that low pitches are lower and high pitches are higher. EXAMPLE 1.6

You can usually find where the stress occurs on a word by trying to tap with your finger in time with each syllable. It is much easier to tap on the stressed syllable. Try saying *abominable* and tapping first on the first syllable, then on the second, then on the third, and so on. If you say the word in your normal way, you will find it easiest to tap on the second syllable. Many people cannot tap on the first syllable without altering their normal pronunciation.

Pitch changes due to variations in laryngeal activity can occur independently of stress changes. They are associated with the rate of vibration of the vocal folds. Earlier in the chapter, we called this the "voice pitch" to distinguish between the characteristic overtones of vowels ("vocal tract pitches") and the rate of vocal fold vibration. Pitch of the voice is what you alter to sing different notes in a song. Because each opening and closing of the vocal folds causes a peak of air pressure in the sound wave, we can estimate the pitch of a sound by observing the rate of occurrence of the peaks in the waveform. To be more exact, we can measure the *frequency* of the sound in this way. **Frequency** is a technical term for an acoustic property of a sound—namely, the number of complete repetitions (cycles) of a pattern of air pressure variation occurring in a second. The unit of frequency measurement is the hertz, usually abbreviated Hz. If the vocal folds make 220 complete opening and closing movements in a second, we say that the frequency of the sound is 220 Hz. The frequency of the vowel [a] shown in Figure 1.4 was 100 Hz, as the vocal fold pulses occurred every 10 ms (one-hundredth of a second).

The **pitch** of a sound is an auditory property that enables a listener to place it on a scale going from low to high, without considering its acoustic properties. In practice, when a speech sound goes up in frequency, it also goes up in pitch. For the most part, at an introductory level of the subject, the pitch of a sound may be equated with its fundamental frequency, and, indeed, some books do not distinguish between the two terms, using *pitch* for both the auditory property and the physical attribute. EXAMPLE 1.7

The pitch pattern in a sentence is known as the **intonation**. Listen to the intonation (the variations in the pitch of the voice) when someone says the sentence *This is my father.* (You can either say the sentences yourself, or listen to the recordings of it in the supplemental materials.) Try to find out which syllable has the highest pitch and which the lowest. In most people's speech, the highest pitch will occur on the first syllable of *father* and the lowest on the second, the last syllable in the sentence. Now observe the pitch changes in the question *Is this your father?* In this sentence, the first syllable of *father* is usually on a lower pitch than the last syllable. In English, it is even possible to change the meaning of a sentence, such as *That's a cat,* from a statement to a question without altering the order of the words. If you substitute a mainly rising for a mainly falling intonation, you will produce a question spoken with an air of astonishment: *That's a cat?*

All the suprasegmental features are characterized by the fact that they must be described in relation to other items in the same utterance. It is the relative values of pitch, length, or degree of stress of an item that are significant. You can stress one syllable as opposed to another irrespective of whether you are shouting or talking softly. Children can also use the same intonation patterns as adults, although their voices have a higher pitch. The absolute values are never linguistically important. But they do, of course, convey information about the speaker's age, sex, emotional state, and attitude toward the topic under discussion.

RECAP

This chapter introduces several key concepts that provide a foundation for later chapters in the book.

- Vocal tract anatomy. We named the major parts of the vocal tract used in speech production.
- Taxonomy of terms describing consonants centering around five aspects of consonant articulation: (1) place of articulation, (2) manner of articulation, (3) voicing, (4) nasality, and (5) laterality.
- Acoustic properties of consonants in waveforms and in spectrograms including amplitude, periodicity, and frequency.
- Taxonomy of vowel description, with three main parameters: front/back, high/low, and rounding.
- Acoustic properties of vowels with a focus on formant frequencies.
- A brief introduction to suprasegmentals, mainly word stress, sentence stress, and intonation.

The organization of this book is cyclical. Each of the concepts introduced in this chapter will appear twice more in the book—once in connection with English phonetics (Part 2) and again in connection with general phonetics (Part 3).