Symbol	Sample Words	Name of Symbol
Syllabic Consonants:		
[m̩]	possum, chasm, Adam, bottomless	syllabic 'm'
[n̩]	button, chicken, lesson, kittenish	syllabic 'n'
[l̩]	little, single, simple, stabilize	syllabic 'l'
[ɹ̩]	ladder, singer, burp, percent	syllabic 'r'
Vowels		
i. Monophthongs (Simple Vowels)		
[i]	beat, we, believe, people, money	
[ɪ]	bit, consist, injury, malignant, business	small capital 'i'
[ɛ]	bet, reception, says, guest	epsilon
[æ]	bat, laugh, anger, comrade, rally	ash
[u]	boot, who, sewer, duty, through	
[ʊ]	put, foot, butcher, could, boogie-woogie	upsilon
[ɔ]	bought, caught, wrong, stalk, core	open 'o'
[ɑ]	pot, father, sergeant, honor, hospital	script 'a'
[ʌ]	but, tough, another, oven	wedge or turned 'v'
[ə]	among, sofa, Asia	schwa
ii. Diphthongs (Complex Vowels)		
[ɑɪ]	bite, Stein, aisle, choir, island	
[ɑʊ]	bout, brown, doubt, flower, loud	
[ɔɪ]	boy, doily, rejoice, perestroika, annoy	
[oʊ]	boat, beau, grow, though, over	
[eɪ]	bait, reign, great, they, gauge	

In the list in the table above, we have given you examples of individual sounds in individual words. When we actually use language on a day-to-day basis, however, we speak in phrases and sentences, with all the words run together. This type of speech is known as **running speech** or **continuous speech**, and, although as linguists we sometimes need to break speech into its component parts of words and sounds, you should bear in mind that most everyday speech is not separated out into these pieces. In running speech, the pronunciations of words may be affected by the surrounding words (see Section 2.2.6 on phonetic co-articulation or File 3.2 on phonological assimilation), and one of the open research questions in the study of language processing is how the human mind processes running speech into its meaningful component parts (see Chapter 9).

FILE 2.2

Articulation: English Consonants

2.2.1 Introducing Articulatory Phonetics

Say the word *hiss* and hold the [s]. Now inhale while holding the tongue position of [s]. What part of your tongue is cooled by the incoming airstream? What part of the roof of your mouth is cooled? Simple, intuitive observations such as these (bolstered by careful X-ray and palatography studies) lead to an **articulatory description** of speech sounds like the consonants of English. **Articulation** is the motion or positioning of some part of the vocal tract (often, but not always, a muscular part like the tongue and/or lips) with respect to some other vocal tract surface in the production of a speech sound (more on this below).

English speech sounds are formed by forcing a stream of air out of the lungs through the oral or nasal cavities, or both. This airstream provides the energy for sound production in the mouth—either by making the vocal folds vibrate or by making hissing or popping noises as air escapes through narrow openings in the mouth. Sounds created by exhaling are said to be made by using a **pulmonic** (=lung) **egressive** (=blowing out) **airstream mechanism**. Other **airstream mechanisms** are used in other languages but are discussed only briefly in Section 2.4.6.

The focus of this file is the articulation of English consonants (refer to Section 2.1.3 for a description of the different types of speech sounds). When describing a consonant it is necessary to provide information about three different aspects of the articulation of the consonant:

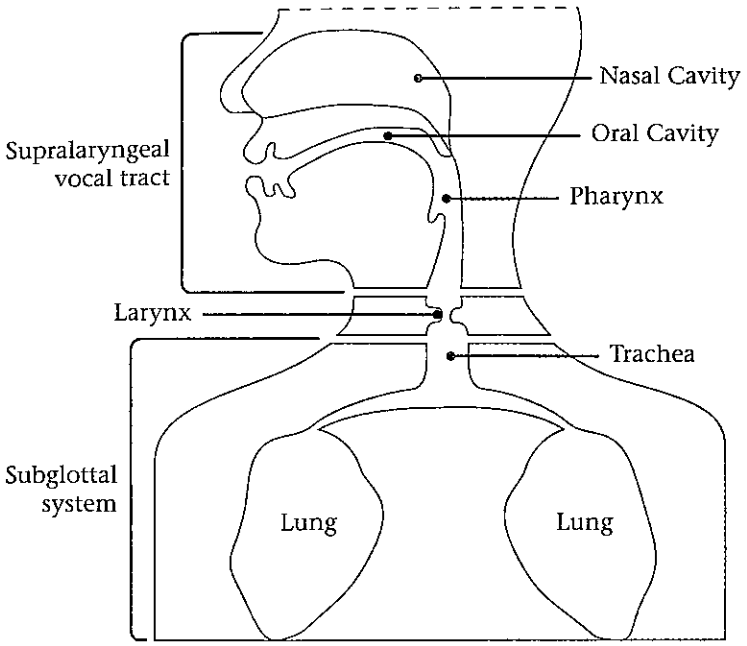
- Is the sound voiced or voiceless?
- Where is the airstream constricted (i.e., the place of articulation)?
- How is the airstream constricted (i.e., the manner of articulation)?

The voicing, place, and manner of articulation are known as **segmental features.** We will discuss each of these aspects in turn. Please note that in this file and elsewhere, whenever we say things like "[p] is voiceless" or "the [p] in *pan,*" what we really mean is "the sound represented by the symbol [p]." Remember that we are talking about speech sounds symbolized by phonetic transcription, but not <p> as in the spelling of the word.

2.2.2 Anatomy of Human Speech Production

In order to answer the three questions listed above, we first need to know more about the anatomy of speech production. There are three basic components of the human anatomy that are involved in the production of speech (see (1)). One is the **larynx** (sometimes called the voice box), which contains the vocal folds and the glottis; another is the **vocal tract** above the larynx, which is composed of the oral and nasal cavities. The third is the **subglottal system,** which is the part of the respiratory system located below the larynx. When air is inhaled, it is channeled through the nasal or oral cavity, or both, through the larynx

(1) The speech production mechanism.



From Lieberman and Blumstein, *Speech Physiology* (1990), p. 4.

into the lungs. English speech sounds are produced while exhaling, as a stream of air is moved out of the lungs and through the larynx and the vocal tract.

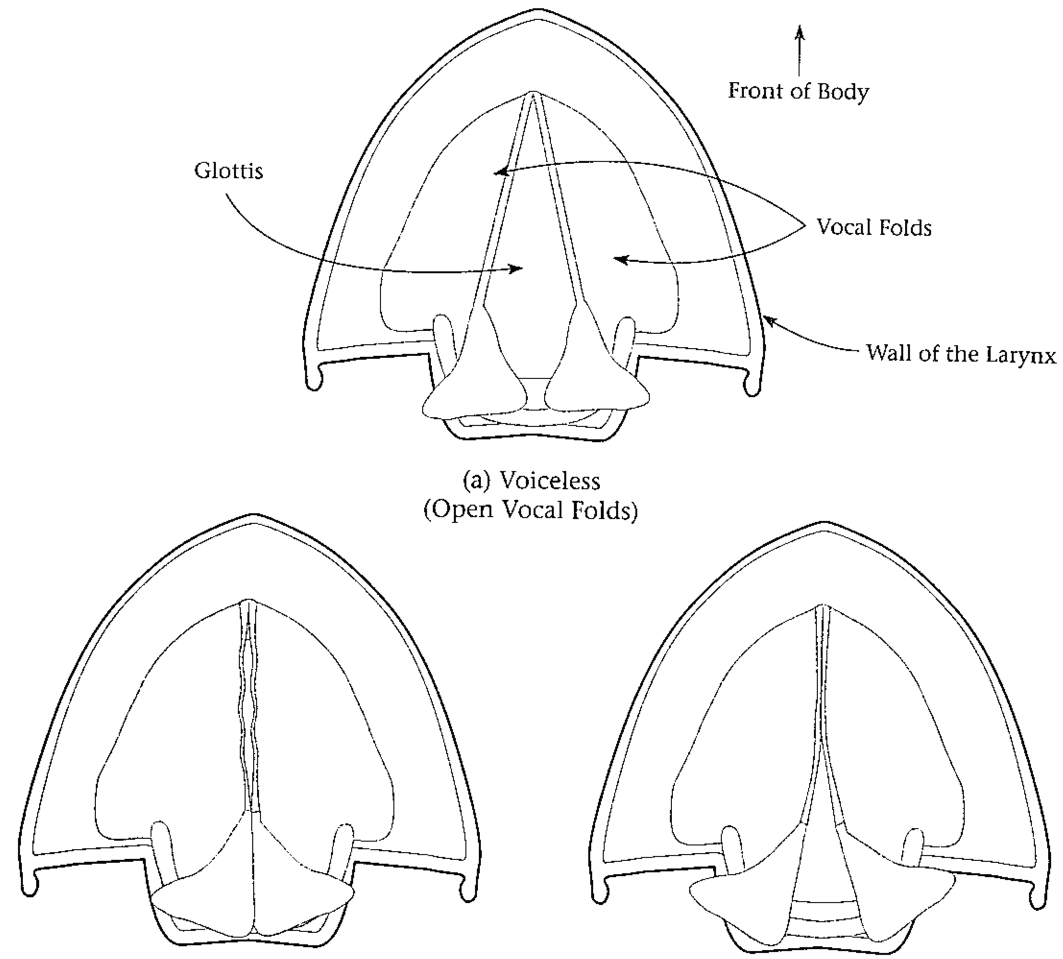
2.2.3 States of the Glottis: Voicing

Humans have a larynx at the top of the **trachea** (or windpipe). Within the larynx are folds of muscle called **vocal folds** (these are popularly known as vocal cords, but they are not really cords). In the display in (2) we are viewing the larynx as if looking down a person's throat. A flap of tissue called the epiglottis is attached at the front of the larynx and can fold down and back to cover and protect the vocal folds, which are stretched horizontally along the open center of the larynx. The opening between these folds is called the **glottis.** At the front of the larynx, the vocal folds are attached to cartilage and can't be moved, but at the back of the larynx, the vocal folds are attached to two small movable cartilages that can close or open the glottis. When the two free ends are brought together, the vocal folds can be nearly or completely closed, impeding airflow through the glottis (2b). When the folds are wide open, the glottis has roughly the shape of a triangle, as can be seen in (2a).

The vocal folds can be opened so that the flow of air coming up from the lungs passes through freely, or the folds may be held close together so that they vibrate as air passes through. Try putting your hand lightly on your throat, or putting your fingers in your ears, and then making a drawn-out [s]. Your vocal folds are separated to open the glottis, as in (2a), so you should feel no vibration. But now make a [z] (again, draw it out), and you will feel a vibration or buzzing feeling. This is due to the vibration of the vocal folds—your glottis is now as in the shape of (2b).

Sounds made with the vocal folds vibrating are called **voiced** sounds, and sounds made without such vibration are called **voiceless** sounds. The underlined sounds in the following pairs of words (see (3)) differ only in that the sound is voiceless in the first word of each pair and voiced in the second. Try saying these words, but don't whisper when you do, because the vocal folds don't vibrate when you whisper.

(2) Three states of the glottis. The view is of the larynx (from above) looking down the throat.



(a) Voiceless
(Open Vocal Folds)

(b) Voiced
(Approximated Vocal Folds)

(c) Whisper
(Partially Closed Vocal Folds)

(3) Voiced versus voiceless sounds

a. [f] fat [v] vat	c. [θ] thigh [ð] thy	e. [s] sip [z] zip	g. [ʃ] dilution [ʒ] delusion
b. [tʃ] rich [dʒ] ridge	d. [p] pat [b] bat	f. [t] tab [d] dab	h. [k] kill [g] gill

In making an articulatory description of a consonant, it is first necessary to state whether a sound is voiced (there is vocal fold vibration) or voiceless (there is no vocal fold vibration). A chart of the voiced and voiceless consonants of English is provided in Section 2.2.7.

Phoneticians can determine if a given segment is voiced or voiceless using a number of different techniques. The simplest is one we described earlier: feeling for vibration of the vocal folds while you produce a sound. This technique, however, is very limited in its ability to determine voicing in running speech (try saying *ice cream* while placing your fingers lightly on your throat—is it obvious that the [s] and [k] in the middle are both voiceless?). One alternative is to examine a picture of the acoustic signal called a **spectrogram,** which will be discussed in more detail in File 2.6. Spectrograms can indicate whether vocal fold vibrations are present in a sound. Another method of studying voicing is to look at the vocal folds directly, using high-speed video. A very thin fiberoptic line is inserted through the speaker's nostril and nasal cavity, down into the upper part of the pharynx. This line conveys a strong white light through the vocal tract to illuminate the vocal folds. A tiny camera, attached to the line and connected to a computer, records movements of the vocal folds. As the subject speaks, the extremely fast vibrations of the vocal folds are filmed so that one can later look at and analyze the recordings frame by frame. While this method allows the speaker to talk freely, with no obstacles in the mouth, and gives a very clear picture of the adjustments and movements of the vocal folds, it is invasive and requires the presence of well-trained medical personnel.

2.2.4 Place of Articulation

In describing consonants, it is also necessary to state where in the vocal tract a constriction is made—that is, where the vocal tract is made narrower. This is referred to as the **place of articulation** of a sound. When reading about each of the following points of articulation, refer to (4), which shows a schematic view of the vocal tract as seen from the side (called a **sagittal section**). To see how this diagram matches up with an actual human head, you may find it helpful to refer to the picture to the lower left, which shows this same diagram superimposed on a photograph.

Bilabial consonants are made by bringing both lips closer together. There are five such sounds in English: [p] *pat,* [b] *bat,* [m] *mat,* [w] *with,* and [w̥] *where* (for some speakers).

Labiodental consonants are made with the lower lip against the upper front teeth. English has two labiodentals: [f] *fat* and [v] *vat.*

Interdentals are made with the tip of the tongue protruding between the front teeth. There are two interdental sounds in most varieties of American English: [θ] *thigh* and [ð] *thy.*

Alveolar sounds are made with the tongue tip at or near the **alveolar** [ælvil̩ɹ] **ridge.** The alveolar ridge is a small ridge that protrudes just behind your upper front teeth. English has seven alveolar consonants: [t] *tab,* [d] *dab,* [s] *sip,* [z] *zip,* [n] *noose,* [l] *loose,* and [ɹ] *red.*

Palatal sounds are made a bit further back in the mouth. If you let your tongue or finger slide back along the roof of your mouth, you will find that the front portion is hard and the back portion is soft. Sounds made with the tongue near the hard part of the roof of the mouth (the "hard **palate**") are called palatal sounds. English makes five sounds in the region of the hard palate: [ʃ] *leash,* [ʒ] *measure,* [tʃ] *church,* [dʒ] *judge,* [j] *yes.* (More precisely, [ʃ, ʒ, tʃ], and [dʒ] are "alveo-palatal" sounds, because they are made in the area between the alveolar ridge and the hard palate. We'll use the shorter term "palatal" to describe these sounds of English, however.)

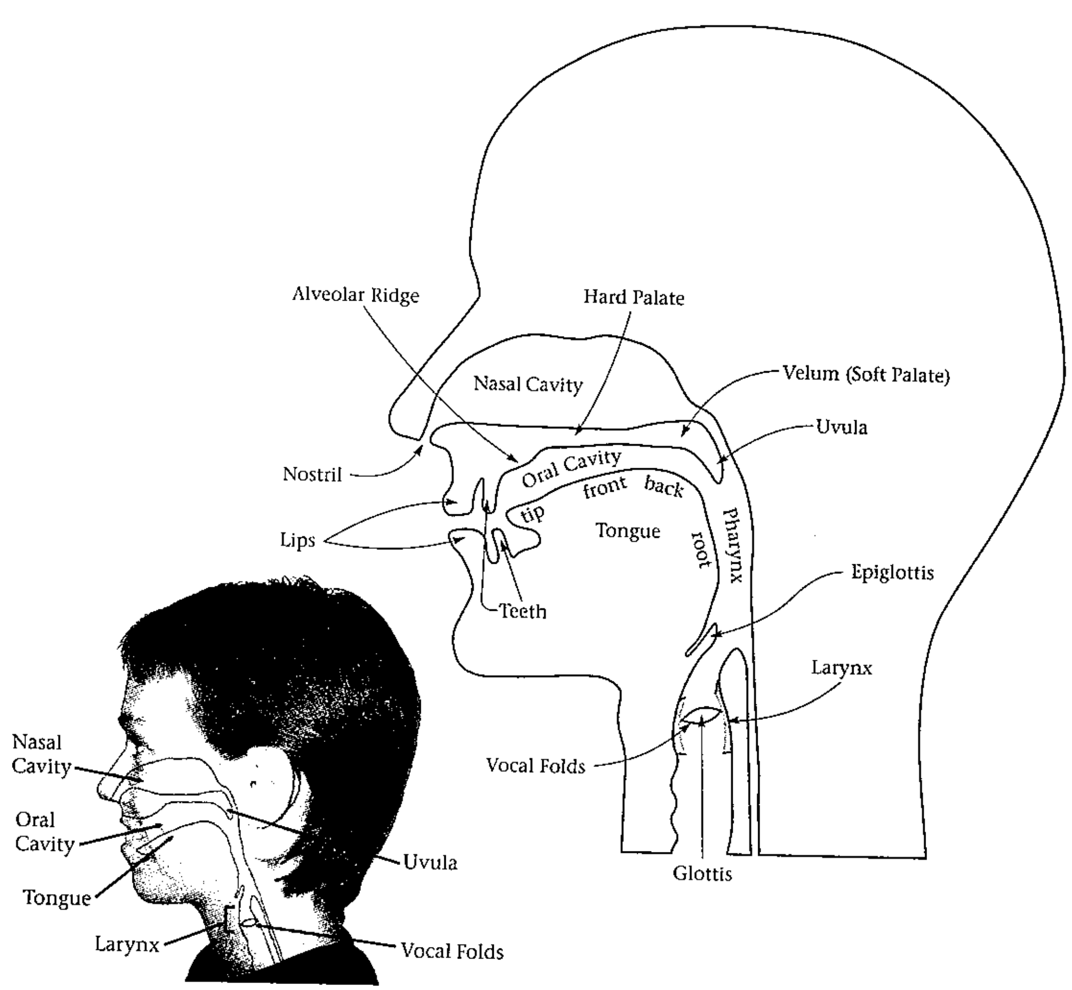
Velar consonants are produced at the soft part of the roof of the mouth behind the hard palate—the **velum.** Sounds made with the tongue near the velum are said to be velar. There are three velar sounds in English: [k] *kill,* [g] *gill,* and [ŋ] *sing.*

Glottal sounds are produced at the larynx. The space between the vocal folds is the glottis. English has two sounds made at the glottis. One is easy to hear: [h], as in *high* and *history.* The other is called a glottal stop and is transcribed phonetically as [ʔ]. This sound occurs before each of the vowel sounds in *uh-oh.*

2.2.5 Manner of Articulation

Besides stating whether a consonant is voiced or voiceless and giving the consonant's point of articulation, it is necessary to describe its **manner of articulation,** that is, how the airstream is modified by the vocal tract to produce the sound. The manner of articulation of a sound depends largely on the degree of closure of the articulators (how close together or far apart they are).

(4) Sagittal section of the vocal tract



Stops are made by obstructing the airstream completely in the oral cavity. Notice that when you say [p] and [b] your lips are closed together for a moment, stopping the airflow. [p] and [b] are bilabial stops. [b] is a voiced bilabial stop, while [p] is a voiceless bilabial stop. [t], [d], [k], and [g] are also stops. What is the three-part description (voicing, place, and manner) of each?

The glottal stop, [ʔ], is made by momentarily closing the vocal folds. If you stop halfway through *uh-oh* and hold the articulators in position for the second half, you should be able to feel yourself making the glottal stop. (It will feel like a catch in your throat.) Nasal consonants are also stops in terms of their oral articulation; see descriptions of nasals below.

Fricatives are made by forming a nearly complete obstruction of the vocal tract. The opening through which the air escapes is very small, and as a result a turbulent noise is produced (much as air escaping from a punctured tire makes a hissing noise). Such a turbulent, hissing mouth noise is called **frication,** hence the name of this class of speech sounds. [ʃ], as in *ship,* is made by almost stopping the air with the tongue near the palate. It is a voiceless palatal fricative. How would you describe each of the following fricatives: [f], [v], [θ], [ð], [s], [z], and [ʒ]?

Affricates are made by briefly stopping the airstream completely and then releasing the articulators slightly so that frication noise is produced. This is why phoneticians describe affricates as a sequence of a stop followed by a fricative. English has only two affricates, [tʃ], as in *church*, and [dʒ], as in *judge*. [tʃ] is pronounced like [t] followed by [ʃ]. It is a voiceless palatal affricate. [dʒ] is a combination of [d] and [ʒ]. What is the three-part description (voicing, place, and manner) of [dʒ]?

Nasals are produced by lowering the velum and thus opening the nasal passage to the vocal tract. When the velum is raised against the back of the throat (also called the **pharynx wall**), no air can escape through the nasal passage. Sounds made with the velum raised are called oral sounds. The sounds [m], as in *Kim*, [n], as in *kin*, and [ŋ], as in *king*, are produced with the velum lowered and hence are called nasal sounds. These consonants are sometimes classified as nasal stops because, just like the oral stops, there is a complete obstruction in the oral cavity. [m] is made with the velum lowered and a complete obstruction of the airstream at the lips. For [n], the velum is lowered and the tongue tip touches the alveolar ridge. [ŋ] is made with the velum lowered and the back of the tongue stopping the airstream in the velar region. In English, all nasals are voiced. Thus [m] is a voiced bilabial nasal (stop); the only difference between [m] and [b] is that the velum is lowered for the articulation of [m], but raised for the articulation of [b]. How would you describe [n] and [ŋ]?

Liquids, like all consonants, involve a substantial constriction of the vocal tract, but the constrictions for liquids are not narrow enough to block the vocal tract or cause turbulence. For the lateral (= side) liquid [l] the midline, or center, of the vocal tract is completely obstructed, like in a stop, but there is a side passage around the tongue. You can feel this positioning by first starting to say *leaf* and "freezing" your tongue at the [l], then inhaling sharply. The air will cool the side(s) of your tongue, showing you the airflow pattern. (Not everyone has the same pattern: do you feel air on the left or right side of your tongue? or both?) The [l] sound is produced with the tongue touching the alveolar ridge as in [t], but the airstream escapes around the sides of the tongue. The presence of this side passage makes [l] sound more like the nasal [n] (which also has a side passage and an alveolar closure) than [t]. [l] is called a lateral liquid. Liquids are usually voiced in English: [l] is a voiced alveolar lateral liquid.

The other liquid in English is [ɹ]. There is a great deal of variation in the ways speakers of English make r-sounds; most are voiced and articulated in the alveolar region, and a common type also involves curling the tip of the tongue back behind the alveolar ridge to make a **retroflex** sound. For our purposes [ɹ] as in *red* may be considered a voiced alveolar retroflex liquid.

Nasals and liquids are classified as consonants, so we would not normally expect them to be syllabic. However, they sometimes act like vowels in that they can function as syllable nuclei. Pronounce the following words out loud, and listen to the liquids and nasals in them: *prism*, *prison*, *table*, and *hiker*. In these words the nucleus of the second syllable consists only of a syllabic nasal or liquid; there is no vowel in these second syllables. In order to indicate that these are **syllabic consonants**, a short vertical line is placed below the phonetic symbol. The final n of *prison* would be transcribed [n̩]; likewise [m̩], [l̩], and [ɹ̩] in *prism*, *table*, and *hiker*.

Glides are made with only a slight closure of the articulators, so that if the vocal tract were any more open, the result would be a vowel sound. [w] is made by raising the back of the tongue toward the velum while rounding the lips at the same time, so it is classified as a voiced bilabial glide. (Notice the similarity in the way you articulate the [w] and the vowel [u] in the word *woo:* the only change is that you open your lips a little more for [u].) [w̥] is produced just like [w], except that it is voiceless; not all speakers of English use this sound. Speakers who use it say it in the word *which* [w̥ɪtʃ], making it distinct from *witch* [wɪtʃ]. [j] is made with a slight closure in the palatal region. It is a voiced palatal glide. Compare the pronunciation of *yawn* [jɔn] and *eon* [iɔn], and notice the similarity between [j] and the vowel [i].

The last manner of articulation that we will discuss here is the **flap.** A flap (sometimes called a tap) is similar to a stop in that it involves the complete obstruction of the oral cavity. The closure, however, is much faster than that of a stop: the articulators strike each other very quickly. In American English, we have an alveolar flap, in which the tip of the tongue is brought up and simply allowed to quickly strike the roof of the mouth as it is returned to its rest position. This voiced sound is symbolized by the IPA character [ɾ] and occurs as the middle sound in the words *writer* and *ladder*.

2.2.6 Investigating Place and Manner of Articulation: Palatography

In addition to the ability of the average speaker to feel at least approximately where and how particular consonant sounds are made, phoneticians have developed a number of methods for looking more precisely at the place and manner of articulation. One of the most common methods is **palatography.** In palatography, a picture is made that shows where the tongue touches the roof of the mouth during a particular articulation.

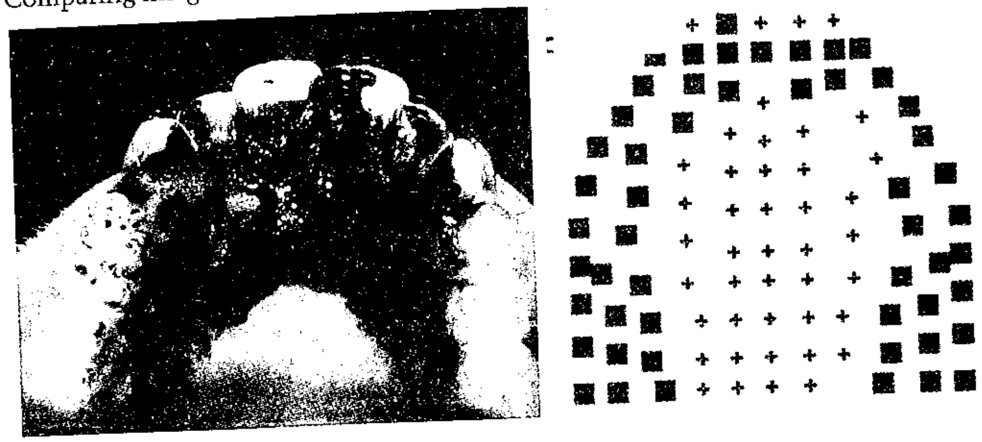
One way to do this, **static palatography,** involves painting the tongue black with a (tasteless) mixture of olive oil and charcoal powder. When the speaker produces a word like *see*, the tongue leaves a black trace on the alveolar ridge and the hard palate, where it touched to make closure. The speaker can then produce another word like *she* (after rinsing off and repainting the tongue), so that the place of articulation for [s] versus [ʃ] can be studied. This method, in addition to being rather messy, works only if the speaker produces a single isolated sound and the contact pattern is photographed or examined immediately.

In order to observe the interplay between articulations, that is, how one consonant's place of articulation affects another consonant's place of articulation, you can use **dynamic palatography** (also called EPG, which is the abbreviation for electropalatography). This method is similar to static palatography but more sophisticated because it allows the experimenter to record sequences of contacts that the tongue makes with the hard palate in the course of the production of an utterance. The places where contact is made are directly recorded into a computer. Once the recordings are made, you can align a specific point in time of the acoustic display of the utterance with a specific EPG display. This way you can measure exactly where, how much, and how long contact between the tongue and the roof of the mouth is produced at any given time in the utterance.

The speaker in such a study is required to use an artificial hard palate (similar to a retainer) that is custom made to fit his or her hard palate exactly. This artificial palate has many small embedded electrodes that record contact as soon as the tongue moves against them. So for any given moment in time during the recording, the researcher knows exactly where the tongue contacts the roof of the mouth. Since the retainer covers only the hard palate, the exact amount of contact made in the soft palate region for velar consonants, such as /g/or /k/, is sometimes hard to see. Nevertheless, this method provides fairly exact data about where and at what point in time within an utterance the tongue touches the hard palate.

You can compare the two types of images made using static versus dynamic palatography in (5). Both of these images show the contact pattern for a [d], spoken by different speakers. The one on the left is the result of static palatography; the one on the right is from dynamic palatography. In both cases, the speaker's teeth are toward the top of the page, and we are looking at the roof of the mouth. In the static palatography picture, the black marks indicate where the tongue touched the roof of the mouth during the production of the nonsense word *ahdah* [ɑdɑ]. In the dynamic palatography picture, the cross-marks indicate the locations of all the sensors on the artificial hard palate; the black boxes indicate sensors that were contacted by the tongue during the [d] of the phrase *bad guy* [bædgɑɪ].

(5) Comparing images from static and dynamic palatography for the production of [d]



In both cases, it is clear that the tongue made full contact with the roof of the mouth, completely closing off the airflow. This is consistent with how we have described the manner of articulation of [d], as a stop.

The contact was made at the front of the mouth in both cases—right along the alveolar ridge, as expected from our description of the place of articulation for [d]. There are differences in the two pictures, however: the one on the left also clearly shows that this speaker produced the word with the tongue far enough forward that it also partly touched the teeth—hence this particular production could be characterized as somewhat **dental**, not purely alveolar. Meanwhile, the speaker on the right has contact that extends much further back along the sides of the mouth. This is very likely because this person was saying the whole phrase *bad guy*, where the [d] is followed by a velar [g]. The two sounds are made using a single flowing action: we don't stop between the two words to reposition the tongue. In this picture, we can see that the back of the tongue is already making contact with the sides of the mouth, in preparation for the velar closure. This kind of **co-articulation** of segments is not something that we can capture using our standard transcriptions, which simply list each segment separately.

Notice that palatography tells you only about the relation of the tongue and the mouth: the pictures in (5) say nothing about the voicing or nasality of the sounds produced. These pictures are completely consistent with the articulations we expect for not only [d] but also [t] (a voiceless alveolar stop) and [n] (a voiced alveolar nasal stop).

2.2.7 The Consonant Chart

The chart of the consonants of English in (6) can be used for easy refeıence. The three-part articulatory description of consonants is conventionally given in this order: Voicing-Place-Manner, e.g., voiced palatal glide or voiceless bilabial stop. To find the description of a sound, first locate the phonetic symbol on the chart. You can find out the state of the glottis by checking whether the sound is in the shaded part of the box or not—the shaded boxes show voiced consonants, while the nonshaded ones show voiceless consonants. Then check the label at the top of the vertical column that contains the sound to see what its place of articulation is. Finally, check the manner of articulation label at the far left of the sound's horizontal row. Locate [ð], for example. It lies in a shaded region, indicating that this sound is voiced. Now look above [ð]. It is in the vertical column marked "interdental." Looking to the far left you see it is a fricative. [ð], then, is the voiced interdental fricative.

You can also use the chart to find a symbol that corresponds to a particular phonetic description by essentially reversing the above procedure. If you want to find the voiced

(6) The consonants of English classified by voicing, place of articulation, and manner of articulation.

Manner of Articulation	Place of Articulation: Bilabial		Labio-dental		Inter-dental		Alveolar		Palatal		Velar		Glottal	
Stop	p	b					t	d			k	g	ʔ	
Fricative			f	v	θ	ð	s	z	ʃ	ʒ			h	
Affricate									tʃ	dʒ				
Flap								ɾ						
Nasal		m						n				ŋ		
Lateral Liquid								l						
Retroflex Liquid								ɹ						
Glide	w̥	w								j				

State of the Glottis: Voiceless | Voiced

palatal fricative, first look in the fricative row, then under the palatal column, and locate the symbol in the shaded part of the box: this is [ʒ].

The chart can also be used to find classes of sounds—that is, groups of sounds that share one or more characteristics. For instance, to find all the alveolars, just read off all the sounds under the "alveolar" column. Or, to find all the stops, read off all the sounds in the "stop" row.

You should familiarize yourself with the chart so that you can easily recognize the phonetic symbols. The list of phonetic symbols for consonants, which was presented in File 2.1.4, should also help you remember which symbol represents which consonant. This chart and the list are also printed on the very last page of this book, for easy reference. Remember that we are talking about *speech sounds* and not letters in the English spelling system.

FILE **2.3**

Articulation: English Vowels

2.3.1 Articulatory Properties of Vowels

In Section 2.1.3, we explained the difference between consonants and vowels, and in File 2.2, we discussed the articulation of consonants. Vowels are the most sonorant, or intense, and the most audible of sounds in speech. Unlike consonants, they usually function as syllable nuclei, and the consonants that surround them often depend on the vowel for their audibility. For example, in the word *pop,* neither [p] has much sound of its own; the [p]s are heard mainly because of the way they affect the beginning and end of the vowel sound.

Because vowels are produced with a relatively open vocal tract, they do not have a consonant-like point of articulation (place of constriction) or manner of articulation (type and degree of constriction). They are also almost always voiced. This means that the three standard descriptors for consonants (place, manner, and voicing) are not helpful when we want to describe vowels. What should we use instead?

Hold your jaw lightly in your hand. Now say *he* [hi], *who* [hu], and *ha* [hɑ]. Did your jaw move for *ha?* The tendency for the jaw to open for [ɑ] is why we will call [ɑ] a **low** vowel. It is usually pronounced with the jaw quite open—lowering the tongue body away from the roof of the mouth. The contrast in jaw position between [i] and [u] as opposed to [ɑ] is large because both [i] and [u] are pronounced with the tongue body close to the roof of the mouth—hence they are called **high** vowels.

Vocal fold vibration is the sound source for vowels. The vocal tract above the glottis acts as an acoustic resonator affecting the sound made by the vocal folds. The shape of this resonator determines the quality of the vowel: [i] versus [u] versus [ɑ], for example.

There are four main ways in which speakers can change the shape of the vocal tract and thus change vowel quality:

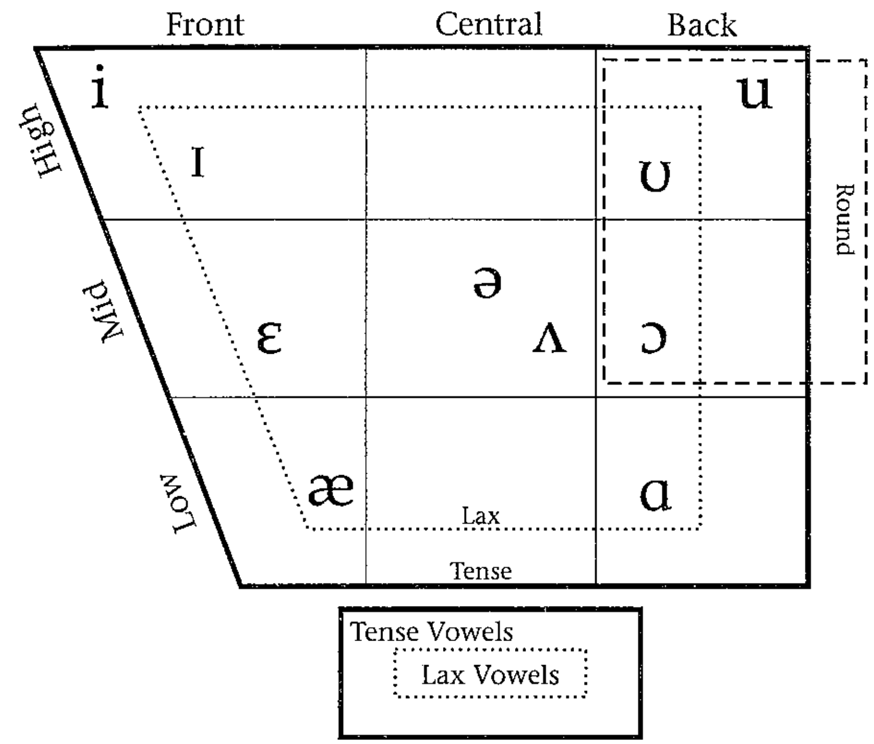
- raising or lowering the body of the tongue
- advancing or retracting the body of the tongue
- rounding or not rounding the lips
- making these movements with a tense or a lax gesture

Therefore, when describing a vowel, it is necessary to provide information about these four aspects of the articulation of the vowel. Refer to the chart in (1) as each aspect is discussed in the following section.[1]

Broadly speaking, there are two types of vowels in English, namely, monophthongs and diphthongs. Diphthongs are two-part vowels, whereas monophthongs have only one part (see Section 2.1.3). We will discuss the four aspects of the articulation of the vowels using monophthongs; articulation of diphthongs will be discussed in the next section.

[1]Although this textbook uses IPA symbols for transcription, the classification of vowels is presented in a more traditional style, with only three levels of height and a tense-lax distinction. The standard IPA vowel chart is printed on the inside back cover of the book for comparison purposes.

(1) The vowels (monophthongs) of English



2.3.2 Tongue Height

If you repeat to yourself the vowel sounds of *seat, set, sat*—transcribed [i], [ɛ], [æ]—you will find that you open your mouth a little wider as you change from [i] to [ɛ], and then a little wider still as you change from [ɛ] to [æ]. These varying degrees of openness correspond to different degrees of tongue height: high for [i], mid for [ɛ], and low for [æ].

High vowels like [i] are made with the front of the mouth less open because the tongue body is raised, or high. The **high** vowels of English are [i], [ɪ], [u], and [ʊ], as in *leak, lick, Luke, look,* respectively. Conversely, **low** vowels like the [æ] in *sat* are pronounced with the front of the mouth open and the tongue lowered. The low vowels of English are [æ] as in *cat* and [ɑ] as in *cot.* **Mid** vowels like the [ɛ] of *set* are produced with an intermediate tongue height. In the inventory of English monophthongs, these mid vowels are [ɛ, ʌ, ɔ], as in *Dell, dull, doll,* respectively. Note that an unstressed vowel in English is often pronounced as the mid vowel [ə], as in *above* and *atomic.*

In many American dialects, words like *caught* and *cot,* or *dawn* and *Don,* are pronounced differently, with an [ɔ] and [ɑ], respectively. In other American dialects, these words are pronounced the same. If you pronounce these pairs the same, you probably use the unrounded vowel [ɑ] in all of these words. For most speakers of English, however, even those who pronounce *caught* and *cot* the same, the vowel [ɔ] appears in words such as *core, more, bore.*

2.3.3 Tongue Advancement

Besides being held high or mid or low, the tongue can also be moved forward or pulled back within the oral cavity. For example, in the high **front** vowel [i] as in *beat,* the body of the tongue is raised and pushed forward so it is just under the hard palate. The high **back** vowel

[u] of *boot*, on the other hand, is made by raising the body of the tongue in the back of the mouth, toward the velum. The tongue is advanced or moved forward for all the front monophthongs, [i], [ɪ], [ɛ], [æ], as in *seek, sick, sec, sack*, and retracted or pulled back for the back monophthongs, [u], [ʊ], [ɔ], [ɑ], as in *ooze, look, hall, dot*. The central vowels, [ʌ] as in *luck* or [ə] as the first vowel in the word *another*, require neither advancement nor retraction of the tongue.

2.3.4 Lip Rounding

Vowel quality also depends on lip position. When you say the [u] in *two*, your lips are **rounded**. For the [i] in *tea*, they are **unrounded**. English has three rounded monophthongs: [u], [ʊ], [ɔ], as in *loop, foot, fall*; all other monophthongs in English are unrounded. In the vowel chart in (1), the rounded vowels are enclosed in a dotted line forming a rectangle.

2.3.5 Tenseness

Vowels that are called **tense** have more extreme positions of the tongue or the lips than vowels that are **lax**. The production of tense vowels involves bigger changes from a mid-central position in the mouth. Additionally, tense vowels in English usually have longer duration (in milliseconds) than lax vowels. On the vowel chart you can clearly see that the distance between the tense vowels [i] and [u] is bigger than the distance between the lax vowels [ɪ] and [ʊ]. For example, tense vowels are made with a more extreme tongue gesture to reach the periphery (outer edges) of the possible **vowel space**. This means that the tongue position for the tense high front vowel [i] is higher and fronter than for the lax high front vowel [ɪ]. Lax vowels, then, are not peripheral to the degree that tense vowels are. Compare tense [i] in *meet* with lax [ɪ] in *mitt*, or tense [u] in *boot* with lax [ʊ] in *put*. In the latter case you will find that the tense rounded vowel [u] is also produced with more and tighter lip rounding than the lax counterpart [ʊ].

Now we can consider some sample descriptions of English vowels:

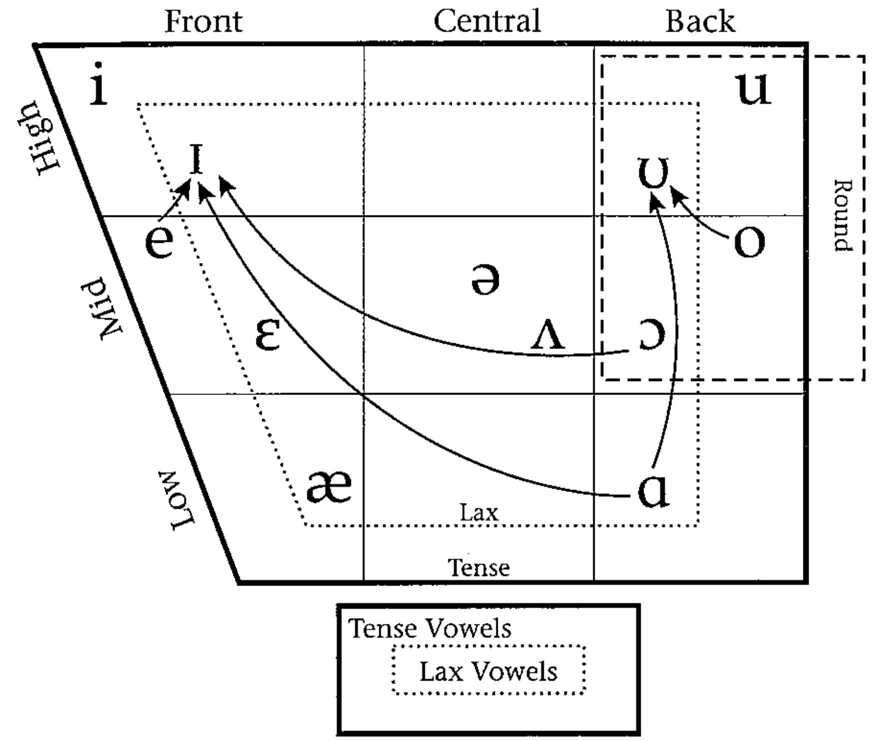
(2) Sample descriptions of English vowels
   a. [i], as in *beat*, is high, front, unrounded, and tense.
   b. [ɔ], as in *caught*, is mid, back, rounded, and lax.
   c. [ɑ], as in *cot*, is low, back, unrounded, and lax.
   d. [ʌ], as in *cut*, is mid, central, unrounded, and lax. (Note that "central" and "mid" refer to the same general area in the vocal tract but along different dimensions.)

2.3.6 Describing Vowels: Diphthongs

You may have noticed that there are a number of vowel sounds in English that have not yet been discussed. For example, the vowels in the words *buy, bay, bow* (as in "bow down"), *bow* (as in "bow and arrow"), and *boy*, or the excalamations *ow!, oh!*, and *oy!* have not yet been described. These vowels are **diphthongs** ([dɪfθɑŋz] or [dɪpθɑŋz]), which, as mentioned in Section 2.1.3, are complex vowel sounds as opposed to monophthongs, which are simple vowel sounds. Diphthongs are "complex" because they are two-part vowel sounds, consisting of a transition from one vowel to the other in the same syllable.

If you try saying the word *eye* slowly, concentrating on how you make this vowel sound, you should find that your tongue starts out in the low back position for [ɑ] and then moves toward the high front position for [ɪ] (see (3)). If you have a hard time perceiving this as two sounds, try laying a finger on your tongue and saying *eye*. This should help you feel the upward tongue movement.

(3) Two-part articulations of the diphthongs of English (the arrows indicate the transitions)



This diphthong, which consists of two articulations and the two corresponding sounds, is written with two symbols: [ɑɪ], as in [bɑɪ] *buy*.[2] To produce the vowel in the word *bow (down)*, the tongue and the lips start in the low back position for [ɑ] and move toward the high back position for [ʊ]; so this diphthong is written [ɑʊ], as in [bɑʊ] *bow (down)*.[2] In the vowel of the word *boy*, the tongue moves from the mid back position for the rounded vowel [ɔ] toward the high front position for [ɪ]; so the diphthong of *boy* is written [ɔɪ], as in [bɔɪ].[3] To say the vowel in the word *bow (and arrow)*, the tongue and the lips start in the mid back position for the rounded vowel [o] and move toward the high back position for the rounded vowel; so the diphthong is written [oʊ], as in [boʊ] *bow (and arrow)*. As for the production of the vowel of the word *bay*, the tongue starts in the mid front position for [e] and moves toward the position for [ɪ]; so this diphthong is written [eɪ], as in [beɪ] *bay*. The chart in (3) illustrates the tongue movements involved in the production of these diphthongs.

2.3.7 Investigating Vowel Articulations

In Section 2.2.6, we described several ways to determine the place and manner of articula-

[2] There are other analyses of the structure of diphthongs. The most common alternative to the one presented here views diphthongs as two-part vowel sounds consisting of a vowel and a glide (see Section 2.2.5) within the same syllable. The correspondence for [ɪ], as in [ɑɪ], is then the palatal glide [j]; hence, [ɑj]. The diphthongs we present as [ɑɪ], [ɑʊ], [ɔɪ], [oʊ], and [eɪ] would be written as [ɑj], [ɑw], [ɔj], [ow], and [ej], respectively, in this system.

[3] We should point out that there is, of course, variation in the pronunciation of all speech, even if we are talking about "Standard American English" (see Chapter 10 on language variation). Some speakers, for example, may produce the vowel in *eye* more like [ai], or the vowel in *bow (down)* more like [əʊ]. If your pronunciations don't exactly match those presented here, you're certainly not wrong! For consistency, however, we will be using the transcriptions given here to represent these sounds.

tion of consonants, using different types of palatography. These methods won't tell us much about vowel articulations, however, because, of course, vowels are produced with a relatively open vocal tract, and the tongue doesn't touch the roof of the mouth. Instead, studying vowels usually involves imaging techniques that allow investigators to look at the whole mouth and the tongue's position in it.

One technique is to use X-ray movies of people talking. These X-ray films can be played over and over again to see tongue, lip, and jaw movements as they occur over time. Although you can find some old example films of X-ray speech by searching online, this methodology is not used anymore because it turned out to be harmful for the speakers.

Instead, researchers now use safer methods such as ultrasound, Magnetic Resonance Imaging (MRI), or Electromagnetic Articulography (EMA). Ultrasound and MRI (like X-rays) both make use of invisible rays that "bounce off" hard structures in their path to create visual images of those structures (in the case of ultrasound, these are sound waves; in the case of MRI, these are radio waves). EMA, on the other hand, involves placing small sensors on a subject's tongue, teeth, and other articulators; these sensors then transmit information back to a computer about their relative locations, allowing researchers to collect precise information about how the articulators move and interact in speech.

Of course, all of the techniques mentioned here can be also used to study consonant articulations, and all are especially useful for consonants that are produced without contact on the hard palate (e.g., [b] or [g]). This makes these techniques particularly well suited for studying the interaction of consonants and vowels in running speech.

FILE 2.4

Beyond English: Speech Sounds of the World's Languages

2.4.1 Beyond English?

In File 2.1, it was claimed that the phonetic alphabet used in this book can be used for any language. The parts of the phonetic alphabet that we have employed up to this point may seem Anglocentric—no different really from *Webster's* pronunciation symbols for English, or any other reasonably consistent method of writing English sounds. To "de-anglicize" our phonetic alphabet so that it is truly useful for describing the pronunciation of other languages, we must add new symbols to it.

It is not the goal of this file, however, to review all of the speech sounds that can be used in human language. Rather, we restrict ourselves to some of the common phonetic symbols that you may encounter. Yet, even this partial look at phonetic diversity highlights the fact that English uses only a small subset of the possible sounds found in human language. We should note that, if you run across a symbol you are not familiar with, you are now in a position to interpret it using the IPA chart on the inside back cover of this book.

2.4.2 Vowels

The most straightforward additions to our phonetic alphabet can be made by filling in some holes. There are certainly other sounds that are possible given the features we've identified for English sounds, but these correspond to combinations of the features that happen not to occur in English. Consider, for example, the vowel chart in File 2.3. In connection with that chart we noted that the only rounded vowels in English are the back vowels [u], [ʊ], and [ɔ] and the diphthong [oʊ] (as in *who'd*, *hood*, *awed*, and *owed*, respectively). You might have thought that these are the only rounded vowels in other languages as well. But if you have studied German or French, you know that this is not true. In addition to the back rounded vowels [u] and [o], German and French both have **front rounded vowels,** such as [y] and [ø]. The high front rounded vowel [y] is pronounced with a tongue position very similar to that for [i], but instead of spread lips, the vowel is pronounced with rounded lips. Similarly, the mid front rounded vowel [ø] is produced with a tongue position as in [e], but with rounded lips. (1) gives some examples of the contrast between front and back rounded vowels in French and in German.

Another vowel distinction that does not come up in English is the distinction between [ɑ] and [a]. [ɑ] is used for low back unrounded vowels, which may contrast with [a], a somewhat more front low unrounded vowel.

All of the vowels we have discussed so far have been oral vowels—that is, they are produced with the velum raised and hence the nasal passage closed. All languages have oral vowels, and many have only oral vowels. Some languages, however, also have **nasalized vowels.**

A nasalized vowel is in nearly every respect identical to its oral vowel counterpart—the only exception is that the nasal passage is open in nasalized vowels (cf. Section 2.2.5 in File 2.2). This is very much like the distinction between an oral stop [b] and a nasal stop [m].

(1) Examples of the contrast between front and back rounded vowels

Front		Back	
		French	
[ty]	'you (familiar)'	[tu]	'all'
[vy]	'seen'	[vu]	'you (formal)'
[nø]	'knot'	[no]	'our (plural)'
[fø]	'fire'	[fo]	'false'
		German	
[gytə]	'benevolence'	[gutə]	'good (masc. sg.)'
[grys]	'greet'	[grus]	'greeting'
[ʃøn]	'beautiful'	[ʃon]	'already'
[bøgen]	'arches'	[bogen]	'arch'

Nasalized vowels are written with a tilde [~] over the corresponding oral vowel symbol. So, a nasalized mid front vowel is written [ẽ], and a nasalized mid back rounded vowel is written [õ].

We don't have to look very far to find vowel nasalization used as the only feature to distinguish words in language, as the following examples from French illustrate (see (2)).

(2) Examples of the contrast between oral and nasal vowels in French

Oral		Nasalized	
[mɛ]	'but'	[mɛ̃]	'hand'
[ʃas]	'hunt'	[ʃɑ̃s]	'luck'
[bo]	'beautiful' (masc.)	[bɔ̃]	'good' (masc.)

2.4.3 Fricatives

Take a look at the fricative row of the English consonant chart in File 2.2. In this row there are five empty cells—bilabial voiceless and voiced, velar voiceless and voiced, and glottal voiced. It turns out that all five of these possible sounds occur in other languages. The symbols that belong in those cells are shown below in (3):

(3) Examples of fricatives

Description	Symbol	Example	Gloss	Language
voiceless bilabial fricative	[ɸ]	éɸá	'he polished'	Ewe
voiced bilabial fricative	[β]	ὲβὲ	'Ewe'	Ewe
voiceless velar fricative	[x]	xɔma	'soil'	Modern Greek
voiced velar fricative	[ɣ]	ɣɔma	'eraser'	Modern Greek
voiced glottal fricative	[ɦ]	pluɦ	'plough'	Ukrainian

Though English does not contrast voiced and voiceless glottal fricatives, we do have the voiced glottal fricative [ɦ] when the *h* sound comes between vowels, as it does in the word *a͟head.*

In theory it should be easy to say the other fricatives in this list because they simply combine features that already exist in English. [ɸ] is a bilabial sound like [p], and a fricative with a noise sounding much like [f]. Voilà, now you can say [ɸ], right? Well, not if you are like most people. It takes practice to master these new, non-English sounds. However, you may have some experience with some of them if you've studied other languages. The voiceless velar fricative [x] is found in German, Yiddish, and Mandarin Chinese. It is the last sound in the German pronunciation of *Bach* [bɑx], the first sound in the Yiddish word [xʊtspə] 'brazenness, utter nerve,' and the first sound in the Mandarin Chinese word [xɑu ^{214}] 'good.' The voiced bilabial fricative [β] is found in Spanish (*Cuba* [kuβa]), as is the voiced velar fricative [ɣ] (*amigo* [amiɣo] 'friend').

2.4.4 Filling in Other Blanks in the Consonant Chart

We can fill in some other empty cells in the English consonant chart. For example, looking at the affricate row, you will notice that English has only palatal affricates. As you might guess, others are possible. For example, the voiceless alveolar affricate [tˢ] occurs in a variety of languages including Canadian French ('ended' [abutˢi]). Similarly, a voiceless labial affricate [pf] is a familar sound from German ('penny' [pfɛnɪk]). The phonetic symbols for these sounds give a good indication of how to say them because we already know how to say [t], [s], [p] and [f].

In addition to the palatal fricatives and affricates of English, it should come as no surprise that some languages make use of palatal stops and nasals. For example, the voiceless palatal stop [c] is used in Greek ('candle' [ceri]), and the voiced palatal nasal [ɲ] is a familiar consonant in Spanish ('pipe' [kaɲa]). These palatal sounds are made with the body of the tongue, like a [k] or [ŋ], but with the tongue touching farther forward in the mouth. You can get the feel of palatal sounds by contrasting your pronunciation of *key,* in which the tongue is fronted, versus *coo,* in which the tongue contact is farther back. It is reasonable to transcribe English *key* as [ci] and *coo* as [ku].

Now, with this description of how to make a palatal stop, we see that the English affricates that we called "palatal" earlier are actually pronounced with a constriction that is somewhere in between alveolar and palatal. This is why we said earlier that [ʃ], [ʒ], [tʃ], and [dʒ] are alveo-palatal consonants. True palatal fricatives do exist, and one example is the voiceless palatal fricative [ç] which is found in Greek ('hand' [çeri]). The five new sounds that we discussed in this section are listed in (4).

(4) Sounds and examples

Description	Symbol	Example	Gloss	Language
voiceless alveolar affricate	[tˢ]	[abutˢi]	'ended'	Canadian French
voiceless labial affricate	[pf]	[pfenik]	'penny'	German
voiceless palatal stop	[c]	[ceri]	'candle'	Modern Greek
voiced palatal nasal stop	[ɲ]	[kaɲa]	'pipe'	Spanish
voiceless palatal fricative	[ç]	[çeri]	'hand'	Modern Greek

2.4.5 Places of Articulation Not Used in English

So far we have seen that the phonetic alphabet contains symbols for non-English sounds that are composed of the same basic phonetic features that are found in English. We now turn to some consonants that are made at places of articulation that we don't find in English.

The voiceless uvular stop [q] is used in Farsi, for example, in the word meaning 'a little bit' [qædri]. The **uvula** is at the very back of the roof of the mouth—that thing that hangs down in your throat. Uvular stops are produced by making a stop closure between the back of the tongue and the uvula. This is like a [k] but with the tongue pulled farther back than normal. The voiced counterpart of [q] is [ɢ].

The voiceless pharyngeal fricative [ħ] is used in Maltese, for example in the word meaning 'clouds' [sħab]. The voiced pharyngeal fricative [ʕ] is used in some dialects of Hebrew, as in the word [ʕor] 'skin.' The pharyngeal place of articulation seems exotic indeed if you thought that the uvular stop had a back tongue position, because the **pharynx** is even further back and lower in the vocal tract. However, it is fairly easy to say a pharyngeal fricative if you start with the vowel [ɑ] of *father* and just open your jaw wider to pull the tongue back in the mouth. For many people this maneuver causes a frication noise—a voiced pharyngeal fricative. The new sounds that we discussed in this section are listed in (5).

(5) Sounds and examples

Description	Symbol	Example	Gloss	Language
voiceless uvular stop	[q]	[qædri]	'little bit'	Farsi
voiceless pharyngeal fricative	[ħ]	[sħab]	'clouds'	Maltese
voiced uvular stop	[ɢ]	[ihipɢeoqteq]	'explore'	Inuktitut
voiced pharyngeal fricative	[ʕ]	[ʕor]	'skin'	Hebrew

2.4.6 Manners of Articulation Not Used in English

Just as some languages use places of articulation that are not used in English, some languages use manners of articulation not found in English. In this section we will describe four non-English manners of articulation.

The American English [ɹ] sound is an exotic speech sound. This sound is very unusual in the languages of the world. It is also very difficult for children to master (e.g., many children pronounce the word *train* as [tweɪn] instead of [tɹeɪn]), and it is also a cause of difficulty for adult learners of English. Most languages that have an /r/ sound have a tongue-tip trilled [r]. If you have studied a language other than English, you may have run into the voiced alveolar **trill** [r]. For example, the sound that corresponds to the Spanish spelling <rr> is trilled ('dog' [pero]).

Another manner of articulation not used in English may be familiar from the Russian word for 'no' [nʲɛt]. The **palatalized** nasal in this word is indicated by the superscript small [ʲ]. To American ears [nʲ] sounds like the sequence [nj], but in X-ray movies of Russian we see that the tongue body position for the glide [j] is simultaneous with the tongue tip position for [n]. So instead of a sequence [nj] the Russian palatalized [nʲ] involves a secondary articulation [ʲ] which is simultaneous with the primary constriction [n]. Many consonants can be palatalized. In the exercises later in this book you will find the palatalized voiceless bilabial stop [pʲ], the palatalized voiceless alveolar stop [tʲ], the palatalized voiceless velar stop [kʲ], the palatalized voiceless alveolar fricative [sʲ], and the palatalized voiceless alveopalatal fricative [ʃʲ].

The phenomenon of secondary articulation helps explain a difference in how [l] is pronounced in English. At the beginnings of words (and as the first sound in stressed syllables within words) [l] is pronounced with the tongue-tip touching the alveolar ridge and the tongue body held rather low in the mouth. But at the ends of words (or as the last sound in a syllable) [l] is pronounced with the tongue body higher in the mouth, and sometimes the tongue-tip does not touch the roof of the mouth at all. Compare the way you say [l] in *laugh* and *Al* (before and after the vowel [æ]). Traditionally these two pronunciations of English [l] are called **clear** (tongue body down, tongue-tip up) and **dark** (tongue body up and tongue-tip down), respectively. We can add to this rough description by noting that in dark [l] (as in *Al*) there is a secondary articulation in which the tongue body moves toward the velum. The dark [l] is therefore more accurately described as **velarized**, and we write this velarized alveolar lateral liquid as [ɫ]. In Macedonian the contrast between velarized [ɫ] and plain [l] distinguishes words: for example, [bela] means 'trouble' while [beɫa] means 'white (fem. nom. sg.).'

The final non-English manner of articulation we want to discuss here is **glottalization.** In glottalized consonants, a glottal stop [ʔ] is produced simultaneously with the primary oral closure in the vocal tract. This simultaneous glottal gesture is symbolized by the small superscript glottal stop symbol [ˀ] after a symbol for whatever consonant is glottalized (e.g., [pˀ] for a glottalized voiceless bilabial stop).

At first, glottalization may seem quite comparable to a secondary articulation. The name for the phenomenon, "glottalization," parallels the names of the secondary articulations "palatalization" and "velarization." Additionally, the symbol used for glottalized consonants (e.g., [pˀ]) is similar in form to the symbols used for palatalized consonants (e.g., [pʲ]).

Unlike palatalization and other secondary articulations, however, glottalization affects the **airstream mechanism** of speech. That is, unlike all of the other sounds we have discussed, the main airstream for glottalized sounds is not the exhaled air from the lungs. Instead, the air pressure that makes the stop release noise (the pop when you release a stop closure) is made by compressing the air in the mouth cavity with the larynx. This is done by closing the glottis (and an oral closure like [k]) and then raising the larynx in the throat. This compresses the air in the mouth—you can think of the rising larynx as a piston in a car engine. Then the stop release noise is made by this compressed air when the [k] closure is released. And then the glottal stop is released. This high-pressure release may make quite a "pop," and so it may not be surprising that such consonants are also called **ejectives.** These consonants may seem very exotic, but they can be fun and easy once you learn them. They occur in 15%–20% of all languages. The sounds that we have discussed in this section are listed in (6).

(6) Sounds and examples

Description	Symbol	Example	Gloss	Language
voiced alveolar trill	[r]	[pero]	'dog'	Spanish
palatalized consonants	[pʲ] etc.	[pʲatʲ]	'five'	Russian
velarized alveolar lateral liquid	[ɫ]	[beɫa]	'white'	Macedonian
glottalized (ejective) stops	[pˀ] etc.	[pˀo]	'foggy'	Lakhota

FILE 2.5

Suprasegmental Features

2.5.1 Segmental vs. Suprasegmental Features

So far we have studied the characteristics of the **segments** (i.e., individual sounds) of speech: place and manner of articulation and voicing for consonants; tongue height and advancement, lip rounding, and tenseness for vowels. In this file we will consider other features that speech sounds may also have: length, intonation, tone, and stress. These features are called **suprasegmental** features because they are thought of as "riding on top of" other segmental features (*supra-* means 'over, above'). Suprasegmental features are different from the features we've studied so far (segmental features) in that it is often difficult or even impossible to identify the quality of a suprasegmental feature if you hear just a single segment. Instead, for suprasegmentals, you have to compare different segments and different utterances to see what the features are. In addition, some suprasegmental features can extend across numerous segments in an utterance, rather than belonging to a single phonetic segment.

2.5.2 Length

The first suprasegmental feature we will talk about is **length**: some speech sounds are longer than others. However, the actual duration of a segment may vary for a number of different reasons (e.g., speaking quickly to a friend as you run out the door versus speaking slowly as you read a story to a young child). Because of this variation, we can't just look at a particular segment and say "that was a long [i]" or "that was a short [i]." Instead, we have to compare the durations of segments within a given utterance (e.g., "this is a long [i] compared to that one").

In some languages, differences in the durations of segments can be as meaningful as the difference between having your tongue body in a high versus a mid front position ([i] versus [e]). Substituting a long segment for an otherwise identical short segment (or vice versa) can result in a different word. For example, consider the data from Finnish shown in (1). In Finnish, both vowels and consonants may be either long or short, and the difference can make a difference in the meaning of a word. (In the data in (1), long vowels and consonants are marked with a following [:]; segments without this symbol are assumed to be short.)

(1) Examples of using length to contrast word meaning in Finnish

a. i. [muta] 'mud'
   ii. [mu:ta] 'some other'
   iii. [mut:a] 'but'

b. i. [tapan] 'I kill'
   ii. [tapa:n] 'I meet'

c. i. [tule] 'come!'
   ii. [tule:] 'comes'
   iii. [tu:le:] 'is windy'

The difference between a long [u:] and a short [u] in Finnish is dependent on the overall speech rate; you have to compare the duration of any given segment with the durations of the other segments to figure out if it was long or short. This is what makes length a suprasegmental feature.

In addition to this type of length difference that can make the difference between two words, speech sounds also vary in duration inherently. For example, all else being equal, high vowels are shorter than low vowels, and voiceless consonants are longer than voiced consonants. Voiceless fricatives are the longest consonants of all.

The duration of a speech sound may also be influence by the sounds around it. For example, say the words *beat* and *bead* aloud. In which word is the [i] longer? In English, a vowel preceding a voiced consonant is about 1.5 times longer than the same vowel before a voiceless consonant. The place and manner of articulation of a following consonant can also affect vowel length. Try saying the word *bees*. How does the length of the [i] in *bees* compare to that in *bead?*

2.5.3 Intonation

Voiced speech sounds, particularly vowels, may be produced with different pitches. Pitch is the psychological correlate of fundamental frequency, which depends on the rate of vibration of the vocal folds (see File 2.6). The pattern of pitch movements across a stretch of speech such as a sentence is commonly known as **intonation.** The intonation contour of an utterance plays a role in determining its meaning. For example, you can read the same words with different intonations and mean different things. Try reading the words in (2) out loud with different pitch patterns, and see if you can get this effect. You might try reading them with either a rising or a falling pitch at the end, or with any other intonation patterns you can think of.

(2) a. You got an A on the test
    b. Yes

Using a rising intonation at the end of the utterance tends to make it sound more like a question, while using a falling intonation makes it sound like a statement.

Although there are multiple systems available for analyzing the intonation of an utterance, one of the most common systems assumes that there are two different intonational phenomena involved in marking the intonation contours of sentences: **pitch accents** and **edge tones.**[1]

Pitch accents usually involve a change in fundamental frequency in the middle of an utterance: a word may be produced with a pitch that is particularly higher or lower than the surrounding words. Words that receive a pitch accent are perceived as very prominent in an utterance—not all words in an utterance get a pitch accent.

Read the examples in (3) aloud. The word that receives a pitch accent, that is, the word that is especially prominent, is written in capital letters. You can see that by putting the prominence on different words, you can use the same string of words to answer different questions.

(3) a. Speaker 1: Who kissed Peter?
    b. Speaker 2: MARY kissed Peter.

    a. Speaker 1: Who did Mary kiss?
    b. Speaker 2: Mary kissed PETER.

[1]This system is know as the ToBI, or "Tones and Break Indices," labeling system. For more information on ToBI, see http://www.ling.ohio-state.edu/~tobi/.

a. Speaker 1: What did Mary do to Peter?
b. Speaker 2: Mary KISSED Peter.

Next we will look at how edge tones can change the meaning of a sentence. Edge tones occur at the end of a phrase. Like pitch accents, they usually involve changes in fundamental frequency, but unlike pitch accents, they represent the pitch pattern right before a perceived break instead of in the middle of an utterance. Read the examples in (4) aloud. In these examples, punctuation indicates where a break occurs. Thus, there is a perceptual break at every period, question mark, and comma. Notice that even though punctuation coincides with a break in these examples, this is not always the case.

(4) a. You got an A on the test.
b. You got an A on the test?
c. You got an A on the test, a C on the homework, and a B on the quiz.

In each example, how did you read the word preceding a punctuation mark? Did you read it with a falling or a rising pitch? The first sentence is a statement and is thus usually produced with falling pitch at the end. This is called sentence-final intonation. The second sentence is a yes/no question, which is usually said with rising pitch, so-called question intonation, at the end. The third example also has sentence-final intonation at the end. But there are two further breaks corresponding to each comma. The pitch before these breaks first falls and then rises again slightly. This is called continuation rise; it indicates that the speaker is not done speaking. Thus, the intonation on the word *test* determines whether the string of words *you got an A on the test* is a statement or a question, and whether the speaker is done speaking or not.

2.5.4 Tone

In many languages, the pitch at which the syllables in a word are pronounced can make a difference in the word's meaning. Such languages are called **tone languages** and include Thai; Mandarin and other "dialects" of Chinese (cf. File 10.1 for an explanation of the notion "dialect"); Vietnamese; languages in New Guinea such as Skou; many of the Bantu languages of Africa such as Zulu, Luganda, and Shona; other African languages such as Yoruba and Igbo; and many North and South American Indian languages such as Apache, Navajo, Kiowa, Mazotec, and Bora. To see how the tone of a word can make a difference in meaning, consider the words in Mandarin Chinese in (5):

(5) Examples from Mandarin Chinese: different tones, different meanings

Segments	Tone Numbers[2]	Tone Pattern	Gloss
[ma]	55	high level	'mother'
[ma]	35	high rising	'hemp'
[ma]	214	low falling rising	'horse'
[ma]	51	high falling	'scold'

As you can see, the same segments in a word (in this case, the syllable [ma]) can be pronounced with different tones and as a result correspond to different meanings.

[2] The tone numbers used in this table were devised by a Chinese linguist named Y. R. Chao to describe the tones of all dialects of Chinese. In this commonly used system for Chinese, '5' indicates the highest pitch and '1' indicates the lowest pitch in the pitch range.

In tone languages, tones can be of two types: either level or contour. All tone languages have level tones; in these tones a syllable is produced with a relatively steady tone such as a high tone, a mid tone, or a low tone. Some tone languages also have contour tones, where a single syllable is produced with tones that glide from one level to another. For example, a rising tone might glide from a low tone to a high tone, while a falling tone might glide from a high tone to a low tone.

There are multiple systems for transcribing tones; the choice of system often has to do with the number and type of tonal contrasts the transcriber needs to make, as well as the history of the systems traditionally used to transcribe tones in a particular set of languages. For Mandarin, for example, tone numbers are often used to indicate the different levels of tone (see (5)). In Kikerewe, on the other hand, tones are often transcribed using accent marks over the vowel in a syllable, where [´] indicates a high tone, [¯] indicates a mid tone, [`] indicates a low tone, [ˇ] indicates a rising tone, and [ˆ] indicates a falling tone (see (6)). See the IPA chart on the inside back cover of the book for the standard IPA symbols used to mark tone.

(6) Examples of level and contour tones in Kikerewe

Word	Tone Pattern	Gloss
[kùsàlà]	low-low-low	'to be insane'
[kùsálà]	low-high-low	'to cut off meat'
[kùʃǐ:ngà]	low-rise-low	'to defeat, win'
[kùsì:ngà]	low-low-low	'to rub, apply ointment'
[kùzúmà]	low-high-low	'to insult, scold'
[kùzùmà]	low-low-low	'to rumble, be startled'
[kùkālâ:ŋgà]	low-mid-fall-low	'to fry'

It is important to note that the tones in a tone language are at least to a certain degree relative, rather than absolute. This is part of what makes them suprasegmental features. For example, the pitch of a high-level tone spoken by a Mandarin speaker with a deep or low-pitched voice will be considerably lower than the pitch of the same tone spoken by a female speaker with a higher-pitched voice. To determine whether a given syllable has a high or a low tone, you must compare it to other syllables spoken by the same speaker—and even then, different utterances may be produced with different tonal ranges!

At the same time, however, there are certain constants in tone production that can help listeners process tones. Some languages tend to be "higher pitched" overall than others: for example, Cantonese tends to be spoken on a higher pitch than Taita, which gives listeners at least some sort of baseline to expect for the tonal range. And, of course, a listener's knowledge about the speaker's physical characteristics (male versus female, tall versus short, etc.) will help him correctly identify the tones he hears.

We should also point out that tone and intonation are not mutually exclusive: tone languages also use intonation.

2.5.5 Stress

The last suprasegmental feature we will examine is **stress**. Stress, like tone, is a property of entire syllables, not just segments. A stressed syllable is more prominent than an unstressed one. This prominence is due to a number of factors, including the fact that stressed syllables are longer and louder than unstressed syllables and usually contain full vowels. Full vowels are produced with more extreme positions of the tongue than reduced vowels, which are produced closer to the mid central position in the mouth and often occur in unstressed syllables.

For example, compare the first vowels in the words *photograph* and *photography;* how are they different? In *photograph,* the first syllable is the most stressed and would be transcribed with the full vowel [oʊ]. But in *photography,* the second syllable is the most stressed, and the vowel in the first syllable has been "reduced" to [ə].

English uses several stress levels, as illustrated by a word like *photography:* in this word, the second syllable is most prominent (has primary stress), the final syllable is next most prominent (has secondary stress), and the other syllables are unstressed (have tertiary stress). In IPA, we transcribe stress using a mark before the beginning of a syllable: primary stress is marked with [ˈ], and secondary stress is marked with [ˌ]. Tertiary stress is not marked. So, for example, the word *photography* would be transcribed as [fəˈtɑgɹəˌfi].

In some languages the placement of stress on a word is predictable; for example, stress almost always falls on the first syllable of a word in Czech, on the next to last syllable of a word in Welsh, and on the last syllable of a phrase in French. In other languages such as Russian and English, stress placement is not predictable and must be learned for each word. In such languages the placement of stress can cause a difference in meaning. For example, what is the difference between a *bláckboard* and a *black bóard*? a *white hóuse* and the *Whíte House*? (Note that in these phrases, an acute accent is placed over the word or syllable that receives primary stress.) Consider also the words *record, perfect,* and *subject.* How are their meanings different when stress falls on the first syllable as opposed to the second? Compare also the words *incite* and *insight,* which differ phonetically only in stress placement but which mean different things.

Much of our emphasis in the previous files has been on the transcription of speech sounds with a series of symbols. Suprasegmental features, on the other hand, prove to be difficult to transcribe this way because they are "superimposed" on the other features. For example, while the symbol [a] always represents the same speech sound whenever we write it, the symbol [ː] has no meaning in isolation. Its meaning is a function of the meaning of the symbol (such as [a]) with which it is used, and even then it indicates only that a segment is long relative to the length of a similar sound transcribed without the [ː]. Similarly, marking stress indicates only that the segments of the stressed syllables are louder and longer than their neighboring sounds. And you can change the intonational pattern of an English utterance radically without changing the segments on which the intonation rides. As you can see, our transcription system doesn't express these facts very well. Perhaps because of this, suprasegmental features remain an important topic in contemporary phonetic research.

FILE **2.6**

Acoustic Phonetics

2.6.1 Articulatory vs. Acoustic Phonetics

So far we have been concerned with articulatory phonetics, the study of how speech sounds are produced. In this file, we will examine many of the exact same speech sounds. This time, however, we will focus on the physical aspects of the sound wave, i.e., the acoustic characteristics of the sounds.

One of the main difficulties in studying speech is that speech is fleeting; as soon as a sound is uttered, it's gone. One of the ways to capture it is to transcribe it using phonetic symbols, as we've seen in previous files. But transcription runs the risk of involving endless debate about what a speaker actually said (e.g., did she say short [a] or long [aː]?). However, modern technology has made it possible to conquer the fleeting nature of speech, at least to some degree, by making records of the acoustic properties of sounds.

2.6.2 Simple Sound Waves

Before we look at speech sounds, it is important to understand something of the nature of sound waves. Sound waves, unlike letters on a page, are not permanent things. They are disturbances in the air set off by a movement of some sort. One kind of movement that can set off a sound wave is vibration, such as that produced by violin strings, rubber bands, and tuning forks—or vocal folds. In this kind of sound wave, a vibrating body sets the molecules of air surrounding it into vibration.

In order to understand how this works, imagine that air molecules are like people in a crowded elevator trying to keep a comfortable distance from one another: if one person moves toward another person, that second person may step back away from the first person. By stepping back, this new person may move closer to yet another person, and so the reaction continues throughout the elevator. Similarly, if one person suddenly moves away from another person, that second person may realize she could have more space on either side by moving back toward the first person. Again, the result may be a chain of movements throughout the crowd while everyone tries to stay equally far apart from everyone else.

There are two physical phenomena resulting from this tendency toward equidistance that make it possible for sound waves to move through the atmosphere. These are **compression,** in which air molecules are more crowded together than usual, and **rarefaction,** in which air molecules are spread farther apart than usual. Because there is a tendency for air molecules (like people in an elevator) to remain equidistant from one another, whenever they are placed in compression or rarefaction, certain instability is set up. Compressed molecules tend to move away from one another so that they are no longer compressed. Likewise, when air is rarefied, there is a tendency for the molecules to move nearer together, as they were before rarefaction occurred.

When the string of a guitar is vibrating, it causes a sound wave in the following way: as the string moves away from its rest position, it pushes the adjacent air molecules closer to neighboring molecules, causing compression. The neighboring, compressed molecules