# FILE 2.0

# What Is Phonetics?

**Phonetics** is the study of the minimal units that make up language.[1] For spoken language, these are the sounds of speech—the consonants, vowels, melodies, and rhythms. As described in File 1.2, the process of communicating has several steps. Within this chain, there are three aspects to the study of speech sounds: **articulatory phonetics,** the study of the production of speech sounds; **acoustic phonetics,** the study of the transmission and the physical properties of speech sounds; and **auditory phonetics,** the study of the perception of speech sounds. In this chapter, we will discuss the articulation and acoustics of speech sounds, as these branches are better understood than auditory phonetics at this point.

One of the most basic aspects of phonetics is figuring out which sounds are possible in speech. You can make a plethora of different noises with your mouth, but only a subset of these noises are used in human language. In this chapter, we will describe some of the features that characterize the speech sounds of the world's languages. We'll see that breaking speech sounds into their component parts reveals similarities among even the most exotic-seeming sounds.

## Contents



[1]While phonetics is traditionally the study of the **sounds** of speech, the study of phonetics is not actually limited to spoken modalities (see File 1.5). Because *phonetics* has come to refer to the study of the minimal units that make up language in general, phoneticians may also study the minimal units (the phonetics) of signed languages (see File 2.7).

FILE **2.1**

# Representing Speech Sounds

## 2.1.1 Studying Pronunciation

"You're not from around here, are you?" Sometimes you can tell by the way a person pronounces words that he or she speaks a dialect that is different from yours. For example, some people do not pronounce *pin* differently from *pen.* In some parts of Ohio the word *push* is pronounced with a vowel sound like the one in *who.* If you hear someone say *poosh,* you can guess where they are from. Such pronunciation differences have been noted for many thousands of years. For example, there is a story in the Bible (Judges 12:4–6) about a group who, after winning a battle, used a password to identify their fleeing attackers. The password they used was *shibboleth,* since their enemies couldn't pronounce the <sh> sound. This group then killed anyone with the telltale pronounciation *sibboleth.* These illustrations show that pronunciation is a part of what we know when we know a language.

There are numerous ways of studying pronunciation in spoken language. In recent years, phoneticians have begun to employ some very sophisticated instrumental techniques to study spoken language.

In articulatory phonetics, we want to know the way in which speech sounds are produced—what parts of the mouth are used and in what sorts of configurations. To investigate these aspects of sound production, phoneticians have used **X-ray photography** and cinematography, among other techniques. More recently, to avoid methods that expose talkers to dangerous amounts of radiation, phoneticians have used point-tracking devices such as the X-ray microbeam or the electromagnetic articulograph to track the locations of small receptors glued onto the lips, tongue, and jaw. Articulatory phonetics is also done with **palatography** (see Section 2.2.6) to observe contact between the tongue and the roof of the mouth, and instruments to measure airflow and air pressure during speech. Ultrasound is also used and is particularly useful for imaging the full tongue during articulation.

In acoustic phonetics, we are more interested in the characteristics of the sounds produced by these articulations. To study acoustic phonetics, phoneticians use pictures of the sounds, using tools such as the **sound spectrograph.** These pictures help acoustic phoneticians explore the physical properties of sounds. These days, you can download sound editing and analysis software from the web. Try searching for a "waveform editor" or an "audio spectrograph," or simply for "phonetics analysis software," and see what free software is available that will to enable you to look at and edit speech sounds on your computer.

The third branch of phonetics, auditory phonetics, focuses on how humans process speech sounds: how we perceive pronunciation. While the fundamentals of perception can be explored by using fairly simple experimental methods that look at human responses to particular stimuli, advanced study of this field depends on more modern equipment such as magnetic resonance imaging (MRI) and computerized tomography (CT).

All of these techniques give us great insight into the details of phonetics. But the simplest and most basic method of phonetic analysis—**impressionistic phonetic transcription**—is still a vital tool for phoneticians. Phonetic transcription is a method of writing down speech sounds in order to capture what is said and how it is pronounced. An example of phonetic transcription is the line "you say tomato, I say tomahto" from Ira Gershwin's lyrics to the song "Let's Call the Whole Thing Off." The word *tomato* is pronounced differently by different people, and we can symbolize two of the pronunciations as "tomato" and "tomahto" as Gershwin did. Or we could follow the pronunciation guide in *Webster's Third New International Dictionary* and write the two pronuncations as tə′mātō and tə′måto. Or we could refer to the *American Heritage Dictionary,* where the two pronunciations are written təmā′tō and təmä′tō. Confusing, isn't it? Yet we need to use phonetic transcription because the normal spelling of the word doesn't tell us enough about how it is pronounced by different people. Spelling is conventionalized, and it symbolizes the word that is meant, rather than how it is pronounced.

| **Spelling** | **Gershwin** | ***Webster's*** | ***Amer. Heritage*** |
|---|---|---|---|
| tomato | tomato | tə′mātō | təmā′tō |
| tomato | tomahto | tə′måtō | təmä′tō |

## 2.1.2 The "Right" Phonetic Alphabet

Did Gershwin write the two pronunciations of *tomato* correctly? Or does one of the dictionaries have the right way to symbolize the difference? It should be clear that there is no one "right" answer about how to write pronunciation in a phonetic transcription. The choices we make are largely arbitrary or influenced by typographical or historical considerations. However, it is absolutely crucial that both the reader and the author agree on the sound qualities that are assigned to the symbols in a phonetic alphabet. This is why almost all dictionaries give some guide to the pronunciation symbols where they list familiar words as examples of the sounds. For example, *fa̲ther* is used to illustrate the sound intended by <å> in *Webster's* and by <ä> in the *American Heritage.* Whether the <a> has one mark or two is an arbitrary decision. This is fine, so long as we have a pronunciation guide.

If the goal of having a phonetic transcription system is to be able to unambiguously convey the important aspects of the pronunciation of a given set of sounds, using a written system of symbols, then such a system must have certain characteristics.

First, each symbol should represent one sound (or **phone**) only, and there should be only one symbol for each sound. The letter <c> violates this principle in English spelling because it represents two sounds (the [k] sound in *cat,* and the [s] sound in *cymbal,* and both the [k] and [s] in *cynic,* for example). Hence, using a <c> does not unambiguously tell the reader which sound is intended.

Second, if two sounds can distinguish one word from another, they should be represented by different symbols. The letters <th> in English violate this principle because the difference between the <th> sounds in *thy* and *thigh* is not captured by using <th> for both words. That is, there is an important difference in pronunciation that is not captured with these letters.

Third, if two sounds are very similar and their difference arises only from the context they are in, we should be able to represent that similarity (see Chapter 3 for how sounds influence surrounding sounds). For example, the [k] sounds in *keep* and *cool* are different from each other in that the exact places they are articulated are dependent on the following vowel. The [k] in *keep* is produced farther forward in the mouth than the [k] in *cool* because the sounds of the words are made using a single flowing action. The influence of one sound on a neighboring sound is known as **co-articulation.** If we are not interested in representing this variation, because it is reasonably predictable in English, we want to make sure that these [k] sounds are not written with different symbols in our transcription system.

Based on the criteria above, the English spelling system is not a good phonetic alphabet because:

- sometimes the same sound is spelled using different letters, such as the [i] sound in *sea, see, scene, receive, thief, amoeba,* and *machine*;
- sometimes the same letters can stand for different sounds, as in *sign, pleasure,* and *resign,* or *charter* and *character,* or *father, all, about, apple, any,* and *age*;
- sometimes a single sound is spelled by a combination of letters, as in *lock, that, book, boast, mountain, shop, apple,* or *special*;
- sometimes a single letter represents a combination of sounds, as in *exit* or *use*;
- sometimes letters stand for no sound at all, as in *know, doubt, though, island, rhubarb,* or *moose.*

A good phonetic transcription system is consistent and unambiguous because there is always a one-to-one correspondence between sounds and symbols. This is even true across languages, so that the symbols you will be learning can be used to transcribe the sounds of any language.

In this book we use the International Phonetic Alphabet (IPA for short). This phonetic alphabet is the right one to use because it is applicable to all spoken human languages, rather than just English, and it has all of the properties of a "useful phonetic alphabet" discussed above. That is, there is a one-to-one correspondence between sounds and symbols, so that each symbol represents only one sound and each sound is represented by only one symbol. In addition, the IPA can be used to transcribe different levels of detail, from broad transcriptions to a very fine level of phonetic detail.

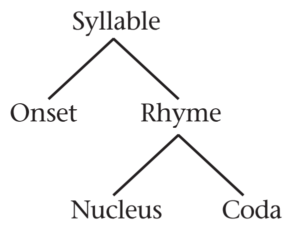
### 2.1.3 Types of Speech Sounds

In order to create a good phonetic transcription system, we need to know what types of sounds we are trying to transcribe. Phoneticians divide the speech stream into two main categories: **segments** and **suprasegmentals**. Segments are the discrete units of the speech stream and can be further subdivided into the categories consonants (File 2.2) and vowels (File 2.3). These sounds are transcribed easily using discrete symbols like [p] and [i]. Suprasegmentals, on the other hand, can be said to "ride on top of" segments in that they often apply to entire strings of consonants and vowels—these are properties such as stress, tone, and intonation (File 2.5). These properties are somewhat more difficult to represent using an alphabetic-like transcription system, and there are many different ways they can be transcribed.

From an articulatory point of view, consonants and vowels are both made by positioning the vocal tract in a particular configuration. However, **consonants** are distinguished from vowels in that consonants are produced with a constriction somewhere in the vocal tract that impedes airflow, while **vowels** have at most only a slight narrowing and allow air to flow freely through the oral cavity. We can also distinguish consonants and vowels acoustically.

Yet another way we can distinguish vowels and consonants is the role each one plays in a **syllable.** A syllable is a unit of speech—every utterance contains at least one syllable. A syllable may contain only a single sound, as in the **monosyllabic** word *uh* [ʌ], or several sounds, as in *sprints* [spɹɪnts]. A syllable can be broken down into an **onset** and a **rhyme.** The rhyme consists of the vowel and any consonants that come after it—the segments that match in words that we think of as rhyming (such as *man, can,* and *plan*)—while any consonants that occur before the rhyme within the syllable form the onset (such as *man, can,* and *plan*). All syllables have a rhyme, but onsets are optional in some languages. The rhyme

(1) Syllable structure




| | | | | |
|---|---|---|---|---|
| e.g., | /k | æ | ts/ | *cats* |
| | /sk | ɪ | n/ | *skin* |

can be further broken down into the **nucleus,** the vocalic part of rhyme, and the **coda,** which consists of any final consonants. The structure of a syllable is shown in (1).

The syllable nucleus is the "heart" of the syllable, carrying suprasegmental information such as stress, volume, and pitch, which vowels are much better suited to do than consonants. Consonants usually do not function as the nucleus of the syllable (but see Section 2.2.5 for syllabic consonants), while vowels do not function as onsets of syllables.

Vowels in turn are often divided into two categories: **monophthongs** ([mɑnəpθɑŋz]) and **diphthongs** ([dɪfθɑŋz] or [dɪpθɑŋz]). You can think of monophthongs as simple vowels, composed of a single configuration of the vocal tract, while diphthongs are complex vowels, composed of a sequence of two different configurations. We consider diphthongs to be "single" vowels, however, because the sequence of two configurations acts as the nucleus to a single syllable. To conceptualize this better, think of the two words *knives* and *naive.* The actual vowel sounds in these two words are essentially the same, but in *knives,* there is just one syllable nucleus (the diphthong [ɑɪ]), while in *naive,* there are two separate syllables with two separate nuclei (the monophthong [ɑ] in the first syllable, followed by the monophthong [i] in the second syllable). The differences between monophthongs and diphthongs will be discussed in more detail in File 2.3.

## 2.1.4 Phonetic Symbols for English

This section lists the IPA symbols for English segments that we will be using in this book. Phonetic symbols are written in square brackets, [ ], to distinguish them from letters or words written in ordinary spelling. It is important to remember that these symbols are not the same as letters of English. Rather, they represent the sounds of language. The following table gives the phonetic symbols for the sound inventory of Standard American English, and the example words make use of Standard American English pronunciations. (Other sounds and symbols will be introduced in File 2.4.) There are recordings of these words available on the Sounds page for Chapter 2.

| Symbol | Sample Words | Name of Symbol |
|---|---|---|
| **Consonants:** | | |
| [p] | pit, tip, spit, hiccough, appear | |
| [b] | ball, globe, amble, brick, bubble | |
| [t] | tag, pat, stick, pterodactyl, stuffed | |
| [d] | dip, card, drop, loved, batted | |

(*cont.*)

| Symbol | Sample Words | Name of Symbol |
|---|---|---|
| [k] | kit, scoot, character, critique, exceed[1] | |
| [g] | guard, bag, finger, designate, Pittsburgh | |
| [ʔ] | uh-oh, hatrack, Batman, button, curtain | glottal stop |
| [f] | foot, laugh, philosophy, coffee, carafe | |
| [v] | vest, dove, gravel, anvil, average | |
| [θ] | through, wrath, thistle, ether, teeth | theta |
| [ð] | the, their, mother, either, teethe | eth, [ɛð] |
| [s] | soap, psychology, packs, descent, peace, excruciating[1] | |
| [z] | zip, roads, kisses, Xerox, design | |
| [ʃ] | shy, mission, nation, glacial, sure | esh, [ɛʃ] |
| [ʒ] | measure, vision, azure, casualty, decision | yogh, [joʊg] or ezh, [ɛʒ] |
| [h] | who, hat, rehash, hole, whole | |
| [ʧ] | choke, match, feature, constituent | |
| [ʤ] | judge, George, Jell-O, region, residual | |
| [m] | moose, lamb, smack, amnesty, ample | |
| [n] | nap, design, snow, know, mnemonic | |
| [ŋ] | lung, think, finger, singer, ankle | engma or eng |
| [l] | leaf, feel, Lloyd, mild, applaud | |
| [ɹ] | reef, fear, Harris, prune, carp | |
| [ɾ] | butter, udder, cuter, Adam, bottle, ready | flap |
| [w] | with, swim, mowing, queen, twilight | |
| [w̥][2] | which, where, what, whale, why (for those dialects in which *witch* and *which* do not sound the same) | voiceless 'w' |
| [j] | you, beautiful, feud, use, yell | lower-case 'j' |
| **Syllabic Consonants:** | | |
| [m̩] | possum, chasm, Adam, bottomless | syllabic 'm' |
| [n̩] | button, chicken, lesson, kittenish | syllabic 'n' |
| [l̩] | little, single, simple, stabilize | syllabic 'l' |
| [ɹ̩][3] | ladder, singer, burp, percent, bird | syllabic 'r' |

[1]The letter <x> in *exceed* and *excruciating* represents a series of two sounds: [ks].
[2]An alternative symbol for this sound is [ʍ], the upside-down 'w.'
[3]Another symbol that is sometimes used for this sound when unstressed (as in *ladder,* but not *burp*) is [ɚ], schwar.

| Symbol | Sample Words | Name of Symbol |
|---|---|---|
| **Vowels:** | | |
| **i. Monophthongs (Simple Vowels)** | | |
| [i] | beat, we, believe, people, money, dean | |
| [ɪ] | bit, consist, injury, malignant, business, gym | small capital 'i' |
| [ɛ] | bet, reception, says, guest, bend | epsilon |
| [æ] | bat, laugh, anger, comrade, rally, hand | ash |
| [u] | boot, who, brewer, duty, through, dune | |
| [ʊ] | put, foot, butcher, could, boogie-woogie | upsilon |
| [ɔ] | bought, caught, wrong, stalk, core, law | open 'o' |
| [ɑ] | pot, father, sergeant, honor, hospital, bomb | script 'a' |
| [ʌ] | but, tough, another, oven, fungus | wedge or turned 'v' |
| [ə] | among, Asia, eloquent, famous, harmony | schwa |
| **ii. Diphthongs (Complex Vowels)** | | |
| [ɑɪ] | I, abide, Stein, aisle, choir, island, fine | |
| [ɑʊ] | bout, brown, doubt, flower, loud | |
| [ɔɪ] | boy, doily, rejoice, perestroika, annoy | |
| [oʊ] | oh, boat, beau, grow, though, over | |
| [eɪ] | bait, reign, great, they, gauge, age | |

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 2.1.2 on phonetic co-articulation or File 3.3 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**, also called an **articulatory gesture**, 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 part of the vocal tract in the production of a speech sound (more on this below).

The focus of this file is the articulation of English consonants. Recall from Section 2.1.3 that consonants are speech sounds that are produced with a constriction somewhere in the vocal tract that impedes the airflow. When describing a consonant, it is therefore necessary to provide information about three different aspects of its articulation:

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

The voicing, place, and manner of articulation are known as **segmental features.** Please remember 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, not letters like <p>, <j>, etc.

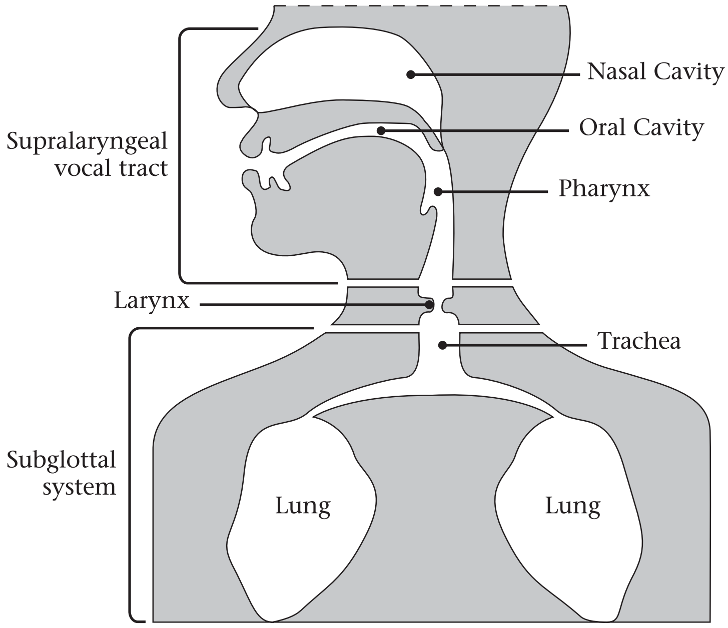
## 2.2.2 Anatomy of Human Speech Production

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)). The first is the **larynx** (sometimes called the voice box), which contains the vocal folds and the glottis and is located in the throat, at the Adam's apple.[1] The second 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, and into the lungs. When air is exhaled, it is forced out of the lungs and through the larynx and the vocal tract.

English speech sounds are formed when exhaling, forcing a stream of air out of the lungs through the oral or nasal cavity, or both. This airstream provides the energy for sound

[1] Yes, everyone has an Adam's apple! It is a protrusion of the cartilage protecting the vocal folds and is usually larger and more prominent in men than in women.

(1) The speech production mechanism.




From Lieberman and Blumstein, *Speech physiology, speech perception, and acoustic phonetics* (1988), p. 4. 

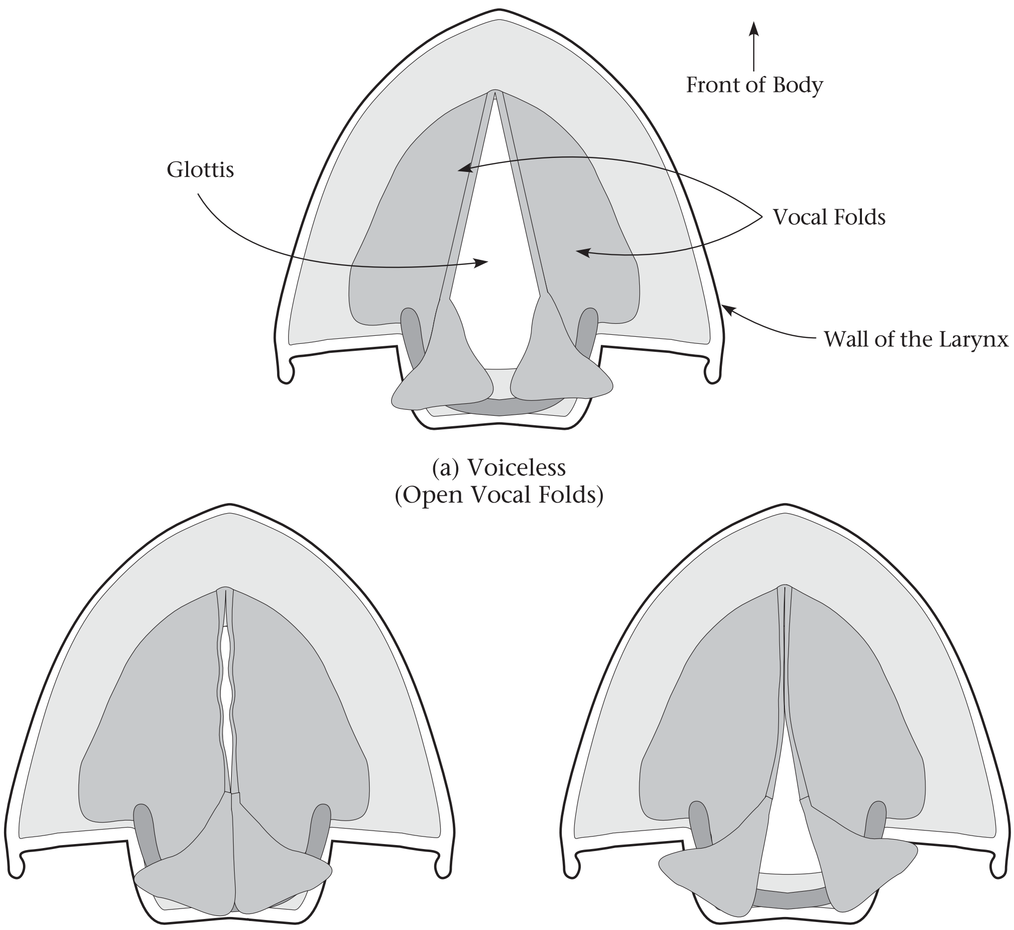
production—either by making the vocal folds vibrate or by making hissing or popping noises as air escapes through narrow openings in the mouth. Sounds created by exhaling are said to be made by using a **pulmonic** (= lung) **egressive** (= blowing out) **airstream mechanism**. All English speech sounds are made using this mechanism, although it is not the only way to produce speech. Other **airstream mechanisms** used in other languages are discussed briefly in Section 2.4.6.

### 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 diagram 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 ("approximated"), 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). There is also an intermediate position, in which the vocal folds are partially open, as shown in (2c). This is the position of the vocal folds when you whisper.

When the vocal folds are open, the flow of air coming up from the lungs passes through freely, and when the folds are held close together, they vibrate as air from the lungs forces its way through them. Try putting your hand lightly on your throat, or putting your

(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)

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). This vibration is called **voicing**.

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.

(3) Voiceless versus voiced sounds

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

In making an articulatory description of a consonant, it is therefore first necessary to state whether a sound is voiced (there is vocal fold vibration; see (2b)) or voiceless (there is no vocal fold vibration; see (2a)). 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. The **voicing bar** on a spectrogram 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

The second aspect of describing consonants is stating where in the vocal tract the 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. We begin our descriptions with the front of the mouth—the left side of the diagram—and work our way back toward the throat.

**Bilabial** consonants are made by bringing both lips close 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 front of the upper **alveolar** [ælvilɹ̩] **ridge.** The alveolar ridges are the bony ridges of the upper and lower jaws that contain the sockets for the teeth. (Think of the inside of a baby's mouth before teeth grow in.) The front of the upper alveolar ridge, which is the most important area in terms of describing alveolar consonants, is the part you can feel protruding just behind your upper front teeth. From now on, any reference to the alveolar ridge means specifically the upper alveolar ridge. English has eight alveolar consonants: [t] *tab,* [d] *dab,* [s] *sip,* [z] *zip,* [n] *noose,* [ɾ] *atom,* [l] *loose,* and [ɹ] *red.*

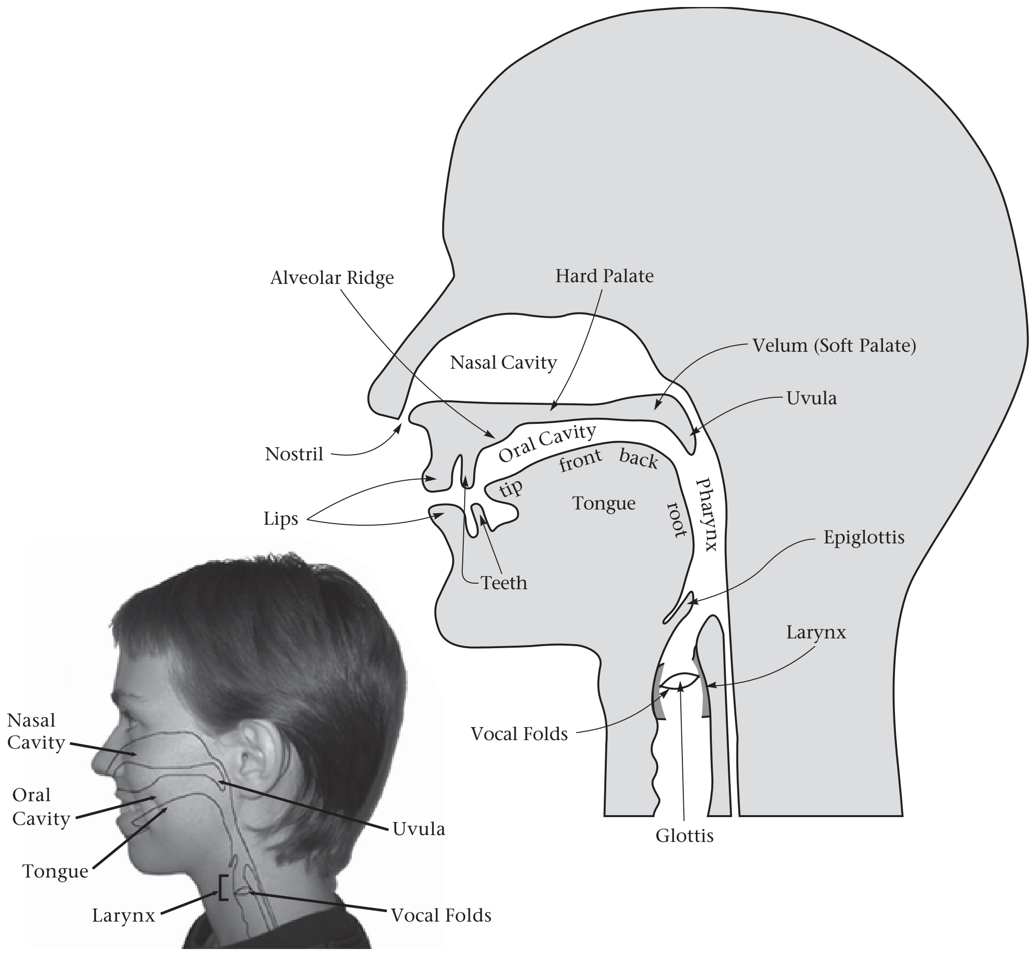
**Post-alveolar** sounds are made a bit farther back. If you let your tongue or finger slide back along the roof of your mouth, you will find that the front portion is hard and the back portion is soft. Post-alveolar sounds are made with the front of the tongue just behind the alveolar ridge, right at the front of the hard palate. English has four post-alveolar sounds: [ʃ] *leash,* [ʒ] *measure,* [ʧ] *church,* and [ʤ] *judge.*

**Palatal** sounds are made with the body of the tongue near the center of the hard portion of the roof of the mouth (the '**hard palate**'). English has only one palatal sound: [j] *y̲es.*

**Velar** consonants are produced at the **velum,** also known as the soft palate, which is the soft part of the roof of the mouth behind the hard palate. Sounds made with the back part of the tongue body raised near the velum are said to be velar. There are three velar sounds in English: [k] *k̲ill,* [g] *g̲ill,* and [ŋ] *sin̲g̲.*

**Glottal** sounds are produced when air is constricted 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 *h̲igh* and *h̲istory*. 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* and in the middle of a word like *cot̲t̲on.*

(4) Sagittal section of the vocal tract




## 2.2.5 Manner of Articulation

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

**Stops** are made by obstructing the airstream completely in the oral cavity. (Stops can also be referred to as plosives, a term that references the release of built-up air pressure when the constriction is opened.) Notice that when you say [p] and [b], your lips are pressed 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 a little different because the closure is not in the oral cavity but in the larynx: the vocal folds momentarily close tight, stopping the airflow. If you press your fingertips lightly on your Adam's apple while saying *uh-oh,* you can feel movement with the closure and then the release in the glottis. And if you stop halfway through *uh-oh,* holding all of your articulators in place, then you should be able to feel a catch in your throat, which is the glottal stop (note that if you keep holding it, you will run out of air, as with all stops!).

**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 just behind the alveolar ridge. It is a voiceless post-alveolar fricative. How would you describe each of the following fricatives: [f], [v], [θ], [ð], [s], [z], [ʒ], and [h]?

**Affricates** are complex sounds, made by briefly stopping the airstream completely and then releasing the articulators slightly so that frication noise is produced. They can thus be described as beginning with a stop and ending with a fricative, as reflected in the phonetic symbols used to represent them. English has only two affricates, [ʧ], as in *church,* and [ʤ], as in *judge.*[2] [ʧ] is pronounced like a very quick combination of a [t], pronounced somewhat farther back in the mouth, followed by [ʃ]. It is a voiceless post-alveolar affricate. [ʤ] is a combination of [d] and [ʒ]. What is its three-part description (voicing, place, and manner)?

**Nasals** are produced by relaxing the velum and lowering it, thus opening the nasal passage to the vocal tract. In most speech sounds, the velum is raised against the back of the throat, blocking off the nasal cavity so that no air can escape through the nose. These sounds are called oral, because the air flows through the oral cavity instead. So when the velum is lowered and air escapes through the nasal cavity, like it is with [m], as in *Kim,* [n], as in *kin,* and [ŋ], as in *king,* the sounds are referred to as nasals. These are sometimes called nasal stops, because there is a complete obstruction of the airflow in the oral cavity, but unlike oral stops, the air continues to flow freely through the nose. For [m], the obstruction is at the lips; for [n], the obstruction is formed by the tongue tip and sides pressing all around the alveolar ridge; and for [ŋ], the obstruction is caused by the back of the tongue body pressing up against the velum. 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]?

**Approximants,** like all consonants, involve constriction of the vocal tract, but the constrictions are not narrow enough to block the vocal tract or cause turbulence. Approximants can be further divided into liquids and glides. We separate these categories mostly because they pattern differently in English, but we will also point out some minor articulatory differences.

**Liquids** are formed with slightly more constriction than glides, and their quality changes (is "liquid") depending on where they occur in a word, e.g., the beginning or end

---

[2]Affricates can also be represented with the two individual symbols that make up the sound, either just next to each other ([tʃ], [dʒ]) or with a tie bar connecting them ([t͡ʃ], [d͡ʒ]). We use the connected versions of the symbols ([ʧ], [ʤ]) in this book to make it clear that affricates function as a single sound.

of a syllable (see the discussion of clear versus dark [l] in Section 2.4.6). The first liquid we have in English is the alveolar lateral liquid [l]. In this sound, the front of the tongue is pressed against the alveolar ridge, as in [d], but unlike in a stop, where the tongue is sealed all the way around the ridge, the sides of the tongue are relaxed (lateral = side), letting the air flow freely over them. You can feel this by starting to say *leaf* and pausing your tongue at the [l], and 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?) Liquids are usually voiced in English, so [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 general alveolar region, and a common type also involves curling the tip of the tongue back behind the alveolar ridge to make a **retroflex** sound. Another common type involves "bunching" the tongue up near the roof of the mouth, but 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. (See Section 2.1.3.) 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 *(o)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 they are fairly close to vowel sounds), and they require some movement (or "gliding") of the articulators during production. [w] is made by raising the back of the tongue toward the velum while rounding the lips at the same time, so it is officially classified as a voiced **labial-velar** glide, though we will usually categorize it as bilabial for the sake of simplicity. (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, for example, the word *which* [w̥ɪʧ], making it distinct from *witch* [wɪʧ]. [j] is made with a slight constriction in the palatal region. It is a voiced palatal glide. Compare the pronunciation of *yawn* [jɔn] and *eon* [iɑn], 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 alveolar ridge before it moves into position for the next sound. 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

The average speaker is able to feel at least approximately where and how particular consonant sounds are made; however, 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 the sound [s] as in *see,* the tongue leaves a black trace wherever it touched to make the constriction.

The speaker can then produce the sound [ʃ] as in *she* (after rinsing off and repainting the tongue), so that the place of articulation (e.g., on the alveolar ridge or the hard palate) 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.

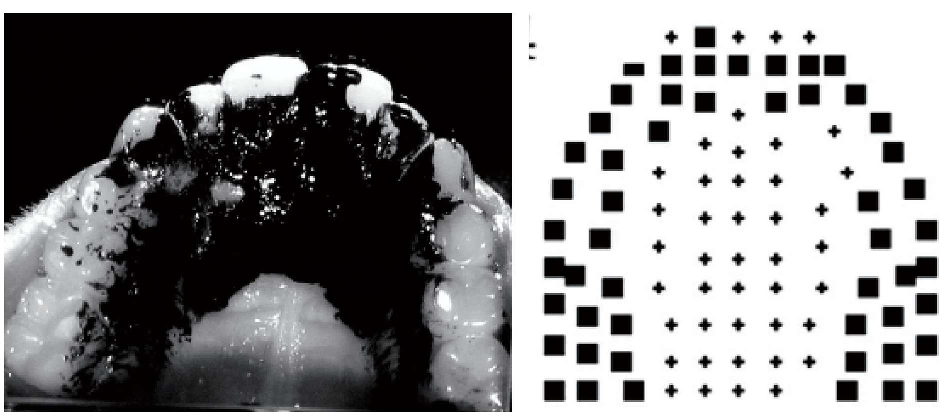
In order to observe the interplay between articulations, that is, how one consonant's place of articulation affects another consonant's place of articulation, you can use **dynamic palatography**. This method is similar to static palatography but more sophisticated because it allows the experimenter 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 dynamic palatography 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. Thus, 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].

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

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



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 also **dental**, not purely alveolar. Meanwhile, the image on the right shows that the speaker's tongue was slightly farther back, on the alveolar ridge but not up against the front teeth. These kinds of minor pronunciation variations are not something that we can capture using our standard transcriptions.

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

### 2.2.7 The Consonant Chart

The chart of the consonants of English in (6) can be used for easy reference. As seen in our descriptions throughout, 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 non-shaded 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 post-alveolar fricative, first look in the fricative row, and then under the post-alveolar 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. (See Section 3.3.2 on natural classes.)

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 last two pages of this book, for easy reference. Remember that we are talking about **speech sounds** and not letters in the English spelling system.

(6) English consonant chart

| Manner of Articulation | Place of Articulation | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Bilabial | | Labio-dental | | Inter-dental | | Alveolar | | Post-Alveolar | | Palatal | | Velar | | Glottal | |
| Stop | p | b | | | | | t | d | | | | | k | g | ʔ | |
| Fricative | | | f | v | θ | ð | s | z | ʃ | ʒ | | | | | h | |
| Affricate | | | | | | | | | ʧ | ʤ | | | | | | |
| Flap | | | | | | | | ɾ | | | | | | | | |
| Nasal | | m | | | | | | n | | | | | | ŋ | | |
| Lateral Liquid | | | | | | | | l | | | | | | | | |
| Retroflex Liquid | | | | | | | | ɹ | | | | | | | | |
| Glide | w̥ | w[3] | | | | | | | | | | j | | | | |

| State of the Glottis | Voiceless | Voiced |
|---|---|---|

[3]As noted above, [w] (along with its voiceless counterpart) is specifically a labial-velar rather than a purely bilabial sound, but we include it in this column for simplicity of representation.

# 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. Here we discuss the articulation of vowels. 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* [pɑp], 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. (See Section 2.6.5 for more information about this.)

Because vowels are produced with a relatively open vocal tract, they do not have a consonant-like place or manner of articulation (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 and the tongue to lie low in the mouth 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:

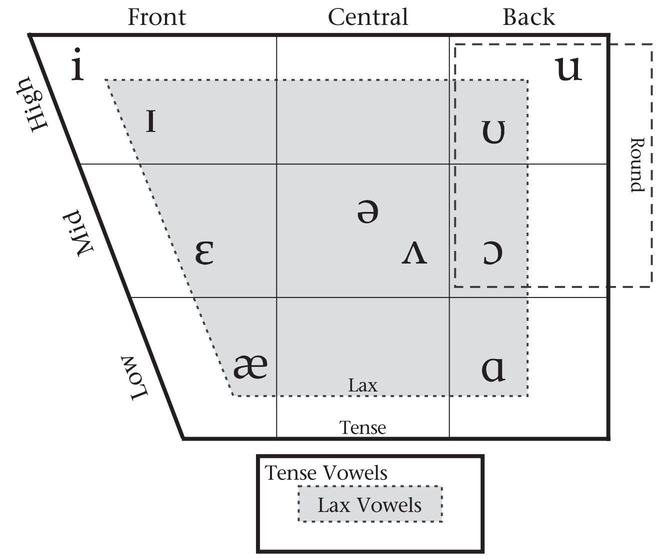
- raising or lowering the body of the tongue
- advancing or retracting the body of the tongue
- **rounding** or not rounding the lips
- making these movements with tense or lax gestures

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]

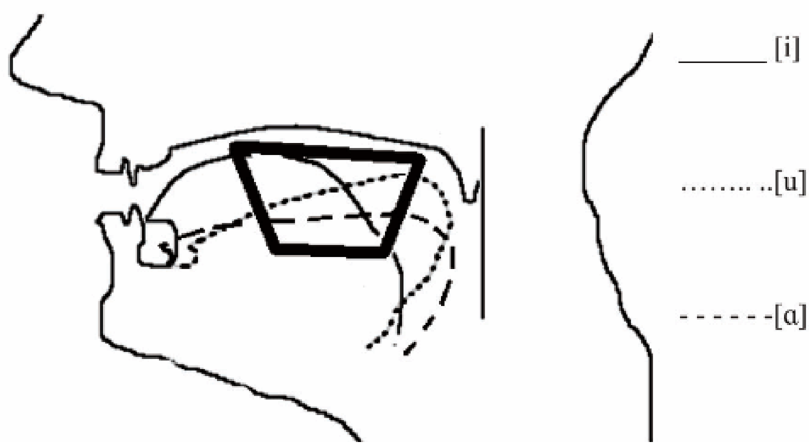
Broadly speaking, there are two types of vowels in English, namely, monophthongs and diphthongs. Diphthongs are two-part vowels, whereas monophthongs have only one part (see Section 2.1.3). We will discuss the four aspects of the articulation of the vowels

[1]Although this textbook uses IPA symbols for transcription, the classification of English 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 last page of the book for comparison purposes.

(1) The vowels (monophthongs) of English




(2) The vowel chart in a sagittal section of the vocal tract, for the articulation of [i], [u], and [ɑ].




using monophthongs; articulation of diphthongs will be discussed in the next section. Note that recordings of many of the following contrasts are available on the Sounds page for Chapter 2, as indicated by the speaker icon.

## 2.3.2 Tongue Height

The first two of these aspects have to do with the position of the tongue in the mouth. Figure (2) shows the position of the tongue for the three vowels discussed above. The highest point of the tongue body is considered to be the place of the vowel, as you can see by comparing the tongue position in (2) with the placement of the symbols in (1).

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.[2] In the inventory of English monophthongs, these mid vowels are [ɛ, ʌ, ɔ], as in *bet, but, bought,* respectively. Note that an unstressed vowel in English is often pronounced as the mid vowel [ə], as in *above* and *atomic*.[3]

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 *law* and *saw,* as well as in words like *core* and *more*.[4]

### 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, as you can see in (2), 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, draw, 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.[5]

---

[2]Some of you may be wondering about the vowel sounds in words like *ate* and *oat*. These sounds begin as the mid front and back tense vowels [e] and [o], respectively. But in most dialects of American English, these vowels are almost always pronounced as the diphthongs [eɪ] and [oʊ], respectively, so they have not been included in the chart of monophthongs. They will be discussed along with the other diphthongs in Section 2.3.6.

[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; vowels in particular vary a lot in English). If at any point your pronunciations don't exactly match those presented here, you're certainly not wrong! For consistency we will be using particular transcriptions to represent sounds throughout the book, but we will make note of specific variations where appropriate. For example, some speakers of American English make further distinctions among unstressed vowels. If you pronounce *roses* differently from *Rosa's,* or the first vowel of *enough* differently from that of *another,* then the underlined vowel in the first word of each pair is likely a high lax vowel, either the front [ɪ] or the central "barred 'i'" [ɨ], while that in the second is the mid [ə] (e.g., [ɪnʌf] vs. [ənʌðɹ]). To keep things simple, we just use [ə] for these unstressed vowels.

[4]Note that vowel characteristics like tense versus lax can be difficult to tell apart before certain consonants in syllable coda position, particularly [ɹ] and [l]. For some speakers, especially those who have a strong contrast between [ɔ] and [ɑ], the vowel in words like *core* and *cold* is closer to a monophthongal [o] than [ɔ], and the vowel in words like *there* is closer to [e] than [ɛ]. We have chosen to use [ɔ] and [ɛ] throughout in these contexts, but the other symbols may well be more appropriate for certain speakers' pronunciations.

[5]You may notice that the vowel symbol [ʌ] is given as back rather than central on the standard IPA chart on the last page of the book. We categorize it as central here based on its pronunciation in most varieties of American English, but it is pronounced as a back vowel in some other varieties, such as New Zealand English.

## 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,** or spread. English has three rounded monophthongs: [u], [ʊ], [ɔ], as in *loop, foot, paw;* all other monophthongs in English are unrounded. In the vowel chart in (1), the rounded vowels are enclosed by the dashed line.

## 2.3.5 Tenseness

Vowels that are called **tense** are said to have more extreme positions of the tongue and/or the lips than vowels that are **lax**. The production of tense vowels involves bigger changes from a mid-central position in the mouth. That is, they require a more extreme tongue gesture of raising, lowering, advancing, or retracting in order 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]. In the vowel chart in (1), 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]. Additionally, tense vowels in English usually have a longer duration (in milliseconds) than lax vowels.

We can use these four characteristics to create the descriptions of English vowels.

(3) Sample descriptions of English vowels
  a. [i], as in *beat,* is high, front, unrounded, and tense.
  b. [ɔ], as in *caught* or *thaw,* 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

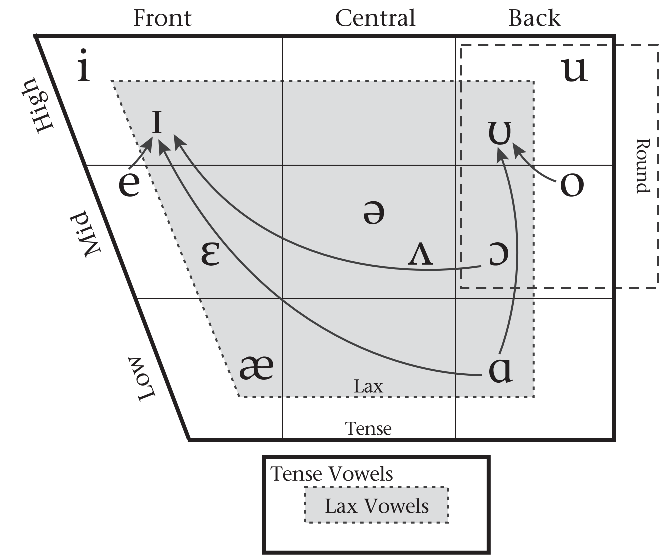
As mentioned in Section 2.1.3, diphthongs are complex vowel sounds, as opposed to monophthongs, which are simple vowel sounds. They are "complex" because they are two-part sounds, consisting of a transition from one vowel to the other in the same syllable. The diphthong in *buy,* which consists of two articulations and two corresponding vowel sounds, is thus written as a two-part symbol: [ɑɪ] as in *buy* [bɑɪ].[6] Although a diphthong consists of a sequence of sounds, it is considered one sound, similar to an affricate (discussed in Section 2.2.5).

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][7] and then moves toward the high front position for [ɪ] (see (4)). If you have a hard time perceiving this

---

[6] 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.

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




as two sounds, try laying a finger on your tongue and saying *eye*. This should help you feel the upward tongue movement.

To produce the vowel in the word *bow (down),* the tongue starts in the low back position for [ɑ] and moves toward the high back position for [ʊ] while the lips round, so this diphthong is written [ɑʊ], as in [bɑʊ]. 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]. To say the vowel in the word *bow (and arrow),* the tongue starts in the mid back position for the rounded vowel [o] and moves toward the high back position for the rounded vowel [ʊ]; so the diphthong is written [oʊ], as in [boʊ]. 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 (4) 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 articulation 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.

[7]There is a great deal of variation in the production of both the low back vowel [ɑ] as in *cot* and the beginning of the diphthongs. While we use the symbol [ɑ] here for both, for many speakers the vowel in *cot* and the first part of the diphthong in *eye* are pronounced differently, and the diphthongs would be more accurately transcribed [aɪ] and [aʊ], representing a more fronted pronunciation of the low vowel. Other variations in diphthong pronunciation include, for example, the production of the vowel in *eye* as more like [ɑi], of the vowel in the word *toy* as more like [oɪ], and of the vowels in *about* and *write* as more like [əʊ] and [əɪ], respectively.

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 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, we said 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 more symbols to it.

It is not the goal of this file, however, to discuss 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 last page 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 more central or front low unrounded vowel. Links to additional examples that you can listen to can be found on our website.

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 velum is lowered and the nasal passage is open, letting air escape through the nose as well as the mouth (cf. Section 2.2.5). This is very much like the

(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' |

distinction between an oral stop [b] and a nasal stop [m]. 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].

While vowels can be nasalized in English, we do not use the nasality to distinguish words. We don't have to look very far, however, to find vowel nasalization used as the only feature to distinguish words in language, as the following examples from French illustrate:

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

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

## 2.4.3 Fricatives

Take a look at the fricative row of the English consonant chart (6) in File 2.2. In this row there are seven empty cells—bilabial voiceless and voiced, palatal voiceless and voiced, velar voiceless and voiced, and glottal voiced. It turns out that all seven 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] | [ɛ̀βɛ̀] | 'Ewe' | Ewe |
| voiceless palatal fricative | [ç] | [çɛri] | 'hand' | Modern Greek |
| voiced palatal fricative | [ʝ] | [ʝut] | 'good' | Dutch |
| 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 *ahead.* We also often pronounce the voiceless fricative at the beginning of words like *huge* as the palatal [ç] rather than the glottal [h]. Both of these pronunciations are due to co-articulation (see Section 2.1.2 and File 3.3).

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 *chutzpah* [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 continue to fill in other empty cells in the English consonant chart. For example, looking at the affricate row, you will notice that English has only post-alveolar affricates. As you might guess, others are possible. For example, the voiceless alveolar affricate [ʦ] occurs in a variety of languages including Canadian French ([abutsi] 'ended'). Similarly, a voiceless labial affricate [pf] is a familar sound from German ([pfɛnɪk] 'penny'). 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 glide [j], 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 ([ceri] 'candle'), and the voiced palatal nasal [ɲ] is a familiar consonant in Spanish ([kaɲa] 'pipe') and Italian ([ɲoki] *gnocchi*). These palatal sounds are made with the body of the tongue, like a [k] or [ŋ], but with the middle of the tongue pressing into the roof of 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 would be reasonable to transcribe English *key* as [ci] and *coo* as [ku]. The four new sounds that we discussed in this section are listed in (4).

(4) Examples of affricates and palatals

| Description | Symbol | Example | Gloss | Language |
|---|---|---|---|---|
| voiceless alveolar affricate | [ʦ] | [abutsi] | 'ended' | Canadian French |
| voiceless labial affricate | [pf] | [pfɛnɪk] | 'penny' | German |
| voiceless palatal stop | [c] | [ceri] | 'candle' | Modern Greek |
| voiceless palatal nasal stop | [ɲ] | [kaɲa] | 'pipe' | Spanish |

## 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 [qædri] 'a little bit.' 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 [sħab] 'clouds.' 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 was pronounced far back in the mouth, 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) Examples of places of articulation not used in English

| **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' | Yemenite 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. Many 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] that 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 post-alveolar 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* (where the [l] is 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,** which produces **ejective** sounds. In ejectives, a glottal stop [ʔ] is produced simultaneously with the primary oral closure in the vocal tract. This simultaneous glottal gesture is symbolized by an apostrophe after the symbol for whatever consonant is glottalized (e.g., [p'] 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." 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 forming 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," giving these consonants the name 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) Examples of manners of articulation not used in English

| Description | Symbol | Example | Gloss | Language |
|---|---|---|---|---|
| voiced alveolar trill | [r] | [pero] | 'dog' | Spanish |
| palatalized consonants | [$p^j$] etc. | [$p^j at^j$] | 'five' | Russian |
| velarized alveolar lateral liquid | [ɫ] | [beɫa] | 'white' | Macedonian |
| ejective consonants | [p'] etc. | [p'o] | '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 segmental features we've studied so far 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 contrast can make a difference in the meaning of a word. Long vowels and consonants are marked with a following [ː]; 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. [muːta] 'some other'
   iii. [mutːa] 'but'

b. i. [tapan] 'I kill'
   ii. [tapaːn] 'I meet'

c. i. [tule] 'come!'
   ii. [tuleː] 'comes'
   iii. [tuːleː] 'is windy'

The difference between a long [uː] 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 contrast 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 influenced 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 **phrase tones.**

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. Pitch accents are therefore used to highlight important information in an utterance, for example.

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.

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

Like pitch accents, phrase tones usually involve changes in fundamental frequency, but unlike pitch accents, they occur at the end of a phrase instead of in the middle of an utterance. Phrase tones have two major functions. First, they affect the overall meaning of an utterance, distinguishing, for example, between a statement, where the speaker provides information, and a question, where the speaker is requesting information. Read the examples in (4a) and (4b) aloud.

(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.

How did you read the last word of each sentence? Did you read it with a falling pitch 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.

Second, phrase tones group words into linguistic units called phrases.[1] A short utterance will usually have just one phrase tone at the end, as in (4a) and (4b), but a longer utterance will usually be broken up into smaller phrases. Read (4c) aloud. Did you notice the perceptual breaks at the commas? (While punctuation coincides with the breaks in these examples, this is not always the case—however, phrase tones can be thought of as the "punctuation" of spoken language.) The pitch before these breaks first falls and then rises again slightly. This is called a continuation rise; it indicates that the speaker is not done speaking. Thus, the intonation on the word *test* does two things: it marks both the end of a phrase and the speaker's intention to continue talking. An example can show how differently a string of syllables will be interpreted depending on the use of phrase tones: say *What's that in the road ahead?* out loud, and then *What's that in the road? A head?* All of the phonetic segments should be exactly the same both times, but a listener will understand the two utterances to have very different meanings!

## 2.5.4 Tone

In many languages, the pitch at which the syllables in a word are pronounced, called the **tone,** 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; 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 Mandarin Chinese words in (5).

[1]Phrases will also be discussed in Chapter 5 on syntax. While the general meaning of the word is the same in both cases—i.e., linguistically relevant groups of words—note that there is not a one-to-one correspondence between the phrases that are marked by phrase tones and syntactic phrases.

(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.

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. These are analogous to diphthongs in that they have two parts. 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. As seen in (5) above, for example, tone numbers are often used to indicate the different levels of tone in Mandarin. In Kikerewe (a Bantu language spoken in Tanzania), on the other hand, tones are often transcribed using accent marks over the vowel in a syllable, where [ˊ] indicates a high tone, [ˉ] indicates a mid tone, [ˋ] indicates a low tone, [ˇ] indicates a rising tone, and [ˆ] indicates a falling tone (see (6)). See the IPA chart on the last page 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' |
| [kùʃǐːngà] | low-rise-low | 'to defeat, win' |
| [kùsìːngà] | low-low-low | 'to rub, apply ointment' |
| [kùzúmà] | low-high-low | 'to insult, scold' |
| [kùzùmà] | low-low-low | 'to rumble, be startled' |
| [kùkālâːŋgà] | low-mid-fall-low | 'to fry' |

It is important to note that the tones in a tone language are at least partially 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! Further, tone and intonation are not mutually exclusive; tone languages also use intonation.

[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.

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 (a Bantu language spoken in Kenya), 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.

## 2.5.5 Stress

The last suprasegmental feature we will examine is **stress.** Stress, like tone, is a property of entire syllables, not segments, though the syllable nucleus, which is usually a vowel, carries most of the information about stress (see File 2.1.3). 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, however, prove to be difficult to transcribe this way because they are "superimposed" on the other features. For example, while the symbol [ɑ] 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 [ɑ]) 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 [ɑ] or long [ɑː]?). 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

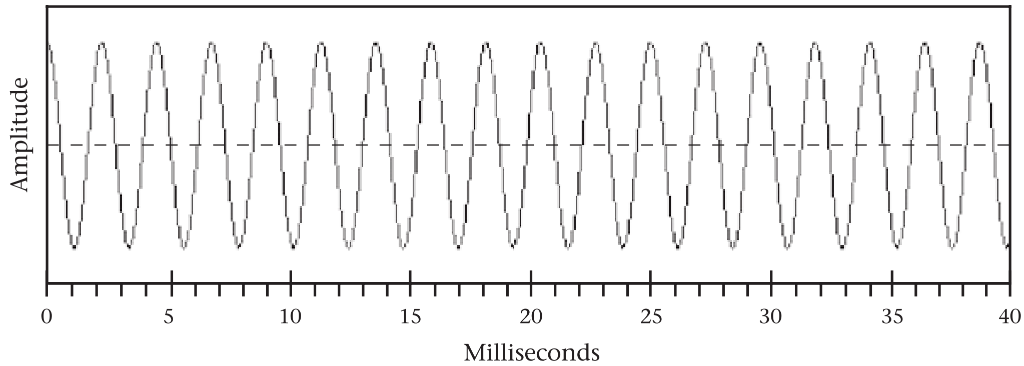
Before we look at speech sounds, it is important to understand something of the nature of sound waves. Sound waves, unlike letters on a page, are not permanent things. They are disturbances in the air set off by a movement of some sort. One kind of movement that can set off a sound wave is vibration, 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 room 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 room. 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 of a tendency for air molecules to remain equidistant from one another, whenever they are placed in compression or rarefaction, a 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. This continues until a stable distance is reached.

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

(1) 440 Hertz sine wave, the tone A




move away from the first "uncomfortably close" molecules, toward others. Those other molecules in turn do the same, and the chain reaction continues.

As the vibrating guitar string moves in the other direction, back to its rest position and beyond, a rarefaction is created. This pulls the air molecules that had been pushed away back toward the string, which creates a rarefaction between them and the molecules on their other side, which pulls those molecules back, and so on. Note that the consequences of the movement (the crowding of the molecules) may be transmitted over a large distance while each individual molecule simply vibrates in place. This chain reaction, which is the consequence of the movement of the string, is the sound wave. When the string moves back and forth at a certain frequency (that is, a certain number of times per second), a group of air molecules that are at some distance from the string will alternately be compressed and rarefied at that frequency. If this chain reaction involving compression and rarefaction is repeated at a rate of 440 times a second, we will hear a musical tone known as "A above middle C." A sound wave such as this, which repeats at regular intervals, is called a **periodic wave.**

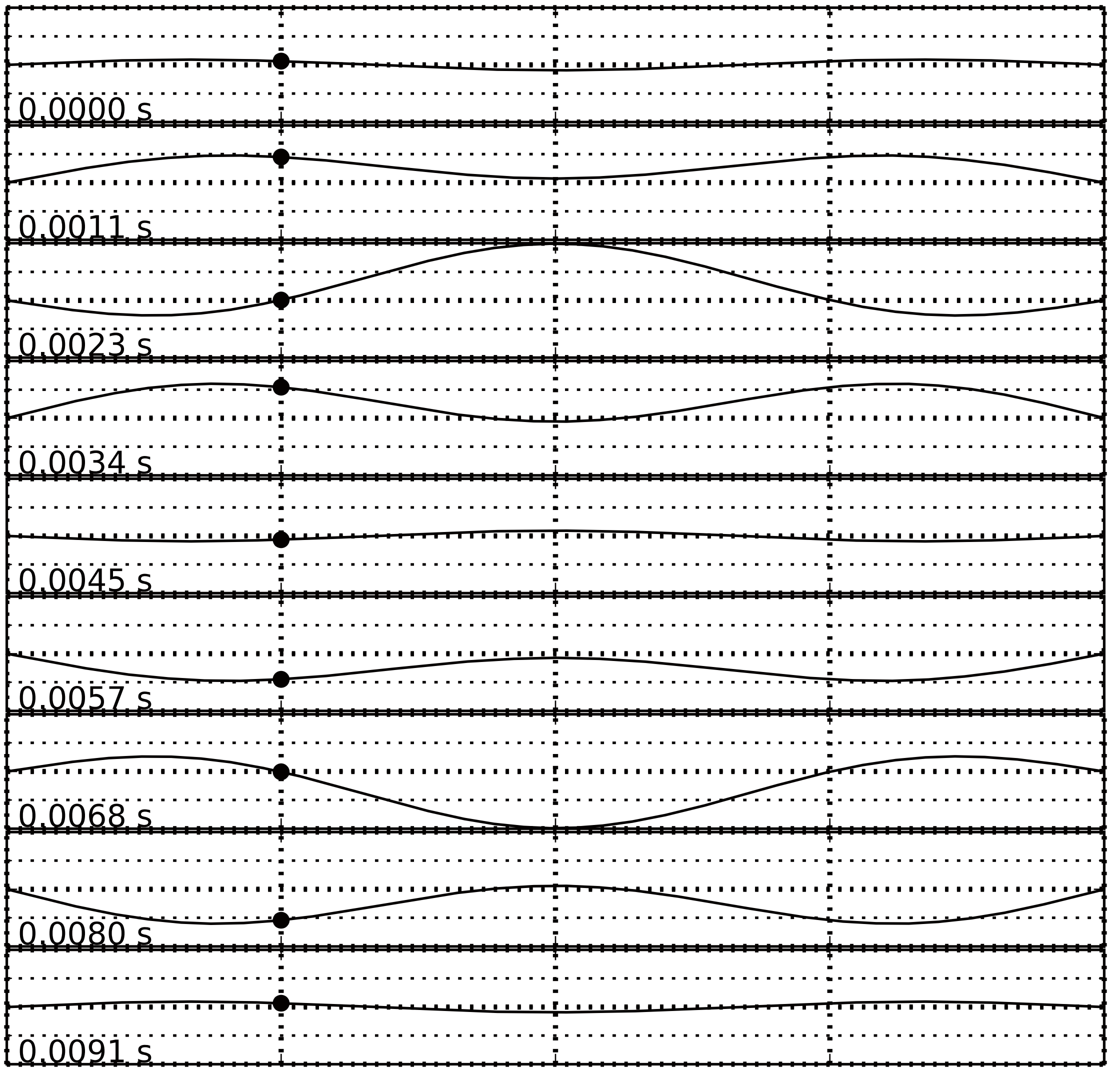
If we plot the energy with which the air molecules press against or pull away from one another in such a sound, the resulting plot looks like the one shown in (1). You can think of the figure in (1) as a plot of the movement (vertical axis) of some air molecules across time (horizontal axis), or, more accurately, you can think of it as being the amount of pressure exerted by the air molecules across time. That is, if the dashed line in the figure represents the resting location of a molecule, you can think of the wavy line (the sine wave) as representing the molecule being pushed away from the resting position, then back toward it, and then away from it in the other direction. The plot in (1) has a frequency of 440 Hz ("Hertz," abbreviated Hz, is a unit of measurement meaning 'cycles/second'), meaning that the molecule moves away from, back toward, away from in the opposite direction, and back toward its original resting position 440 times in a single second.

Air molecules can vibrate at many different frequencies. When they vibrate at rates from 20 to 20,000 times a second, we perceive the vibration as sound. It is interesting to note, however, that we don't really use this whole range for speech. In fact, the highest frequency that can be transmitted by a telephone is 3500 Hz, and yet little essential information about the speech signal is lost by cutting off frequencies above this. And, of course, air molecules vibrate at higher or lower frequencies that we do not hear as sound at all. (But some animals do hear them as sound!)

### 2.6.3 Complex Sound Waves

Our discussion of sound waves up to this point has been very basic and somewhat simplified. In fact, simple sound waves such as those discussed in the previous section are not

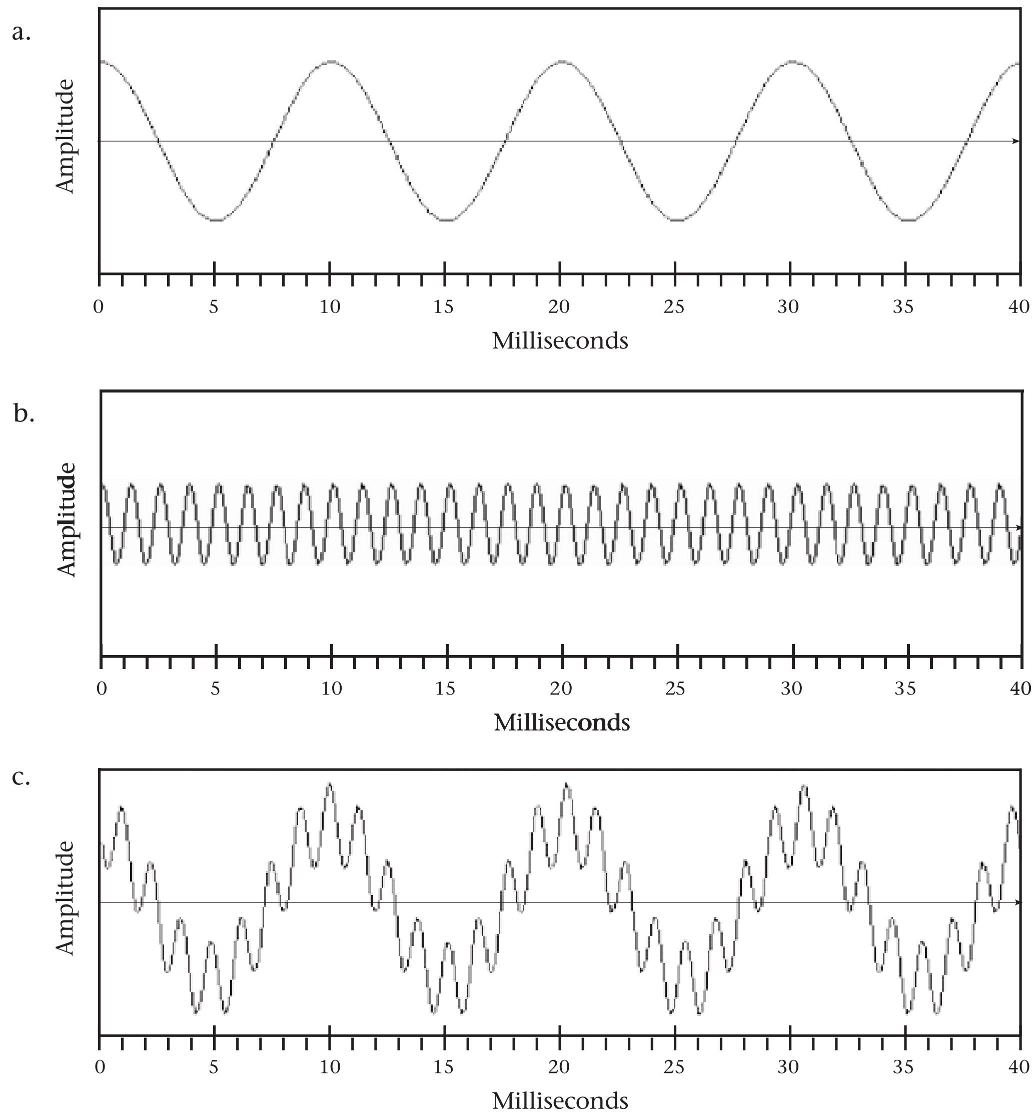
(2) A point on a guitar A string




necessarily produced by guitar strings or human vocal folds. Most things can vibrate at many frequencies at the same time. Figure (2) shows a particular point of a guitar A string: moving up, then part of the way back down, then back up again, then all the way down, then part of the way back up, then back down again, and finally all the way up again to start over. This pattern is repeated 110 times per second. But, since the motion of the string is more complex than just going all the way up and all the way back down, you can describe the string as vibrating both at 110 Hz (the number of times per second the entire pattern repeats) and at 330 Hz (the number of times per second a partial pattern is repeated.

Very complex waves can result from this sort of combination of movements, but the complex movements can always be broken down into simple movements at simple frequencies. These simple movements at single frequencies in turn generate simple sound

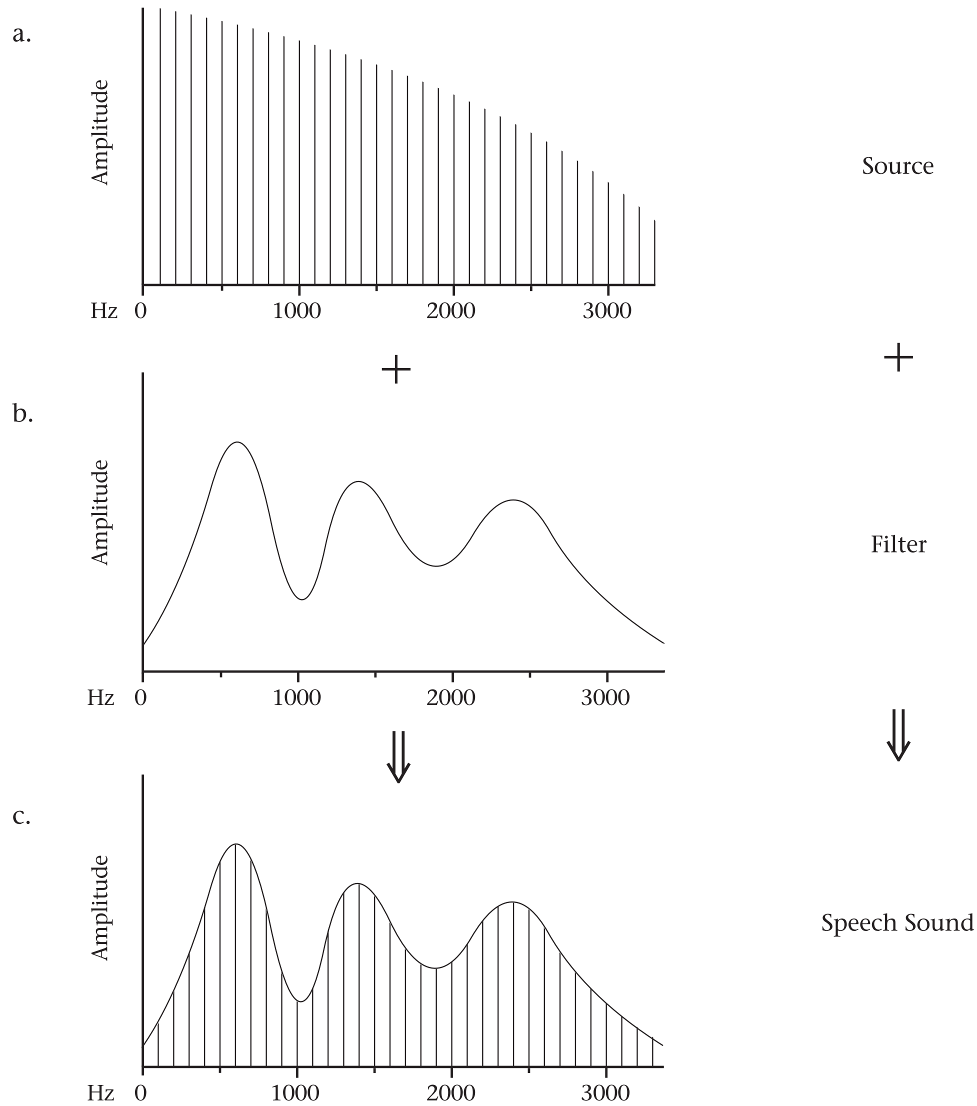
(3) Two simple waves combining to form a complex wave




waves at single frequencies that combine to make complex sound waves. For example, in Figure (3), you can see a plot of two simple sound waves, (a) and (b), and the complex sound wave (c) resulting from their combination. Thus (3a) and (3b) illustrate the simple wave components of the complex wave in (3c).

The sound wave that is produced by the vocal folds is a complex wave. This complex wave is composed of a fundamental wave, which repeats itself at the frequency of the opening and closing of the vocal folds, and a set of **harmonics,** which repeat at frequencies that are multiples of the fundamental. Thus, if the vocal folds open and close at a rate of 100 cycles per second, the fundamental frequency of the resulting sound wave is 100 Hertz (cycles/second), the second harmonic is 200 Hz, the third harmonic is 300 Hz, and so on. Note that the first harmonic is the **fundamental frequency** (pitch).

(4) Source plus filter equals speech sound




The complex wave produced by the vocal folds is known as the source wave, because the vocal folds are the source of the sound wave: it is their movement that creates the wave. It can be represented in a histogram as in (4a), where the horizontal axis represents frequency, and the vertical axis represents the amplitude of the wave. Each line represents one component wave (or harmonic) in the complex vocal wave. Note that the relative amplitude of each wave gets progressively smaller at higher frequencies.

As this sound wave passes through the vocal tract, the articulators shape it, or filter it, boosting the energy at some harmonic frequencies and damping the energy at others. This filter action is similar to the effect of room acoustics on a speaker's voice. Some rooms enhance the voice so that no amplification is needed, while others seem to absorb the voice, muffling the sound. In a similar way, the vocal tract acts as a filter on the source wave. In (4), the vocal tract positioned for the vowel [ɑ] has a filtering effect as in (4b), and harmonics at about 600 Hz, 1380 Hz, and 2500 Hz are enhanced, while harmonics at other positions are damped, yielding the output wave in (4c).

Thus a speech sound (wave) is the result of two independent things: the source wave (the contribution of the vocal folds) and the filter (the contribution of the articulators and the vocal tract).

## 2.6.4 Vowels

In the production of vowels, the filtering effect of the vocal tract produces amplitude peaks at certain frequencies by enhancing the harmonics (the component waves of a complex waveform, produced by the vocal folds) at those frequencies while damping harmonics at other frequencies, as discussed above. These peaks in the filter function are called **formants** (resonant frequencies of the vocal tract). For example, just as a trombone has particular resonant frequencies (determined by the length of the tube) that shape the sound produced by the vibration of the lips, in vowel sounds the vocal tract has resonant frequencies (determined by the length and configuration of the vocal tract) that shape the sound produced by vocal fold vibration. Vowels have several formants, the first three of which are the most important for speech perception (see File 9.4 for a discussion of speech perception). The values of these formants differ from vowel to vowel, which allows us to distinguish between vowels we hear. The table in (5) lists typical formant frequencies for eight American English vowels.

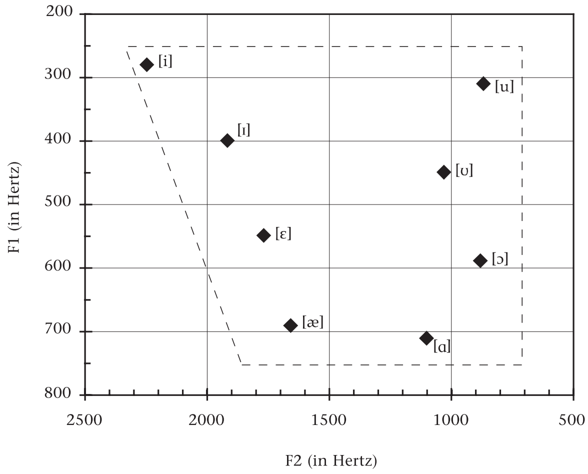
(5) Typical frequencies in Hz of the first, second, and third formants for American English vowels

| Vowel | F1 | F2 | F3 |
|---|---|---|---|
| [i] | 280 | 2250 | 2890 |
| [ɪ] | 400 | 1920 | 2560 |
| [ɛ] | 550 | 1770 | 2490 |
| [æ] | 690 | 1660 | 2490 |
| [u] | 310 | 870 | 2250 |
| [ʊ] | 450 | 1030 | 2380 |
| [ɔ] | 590 | 880 | 2540 |
| [ɑ] | 710 | 1100 | 2540 |

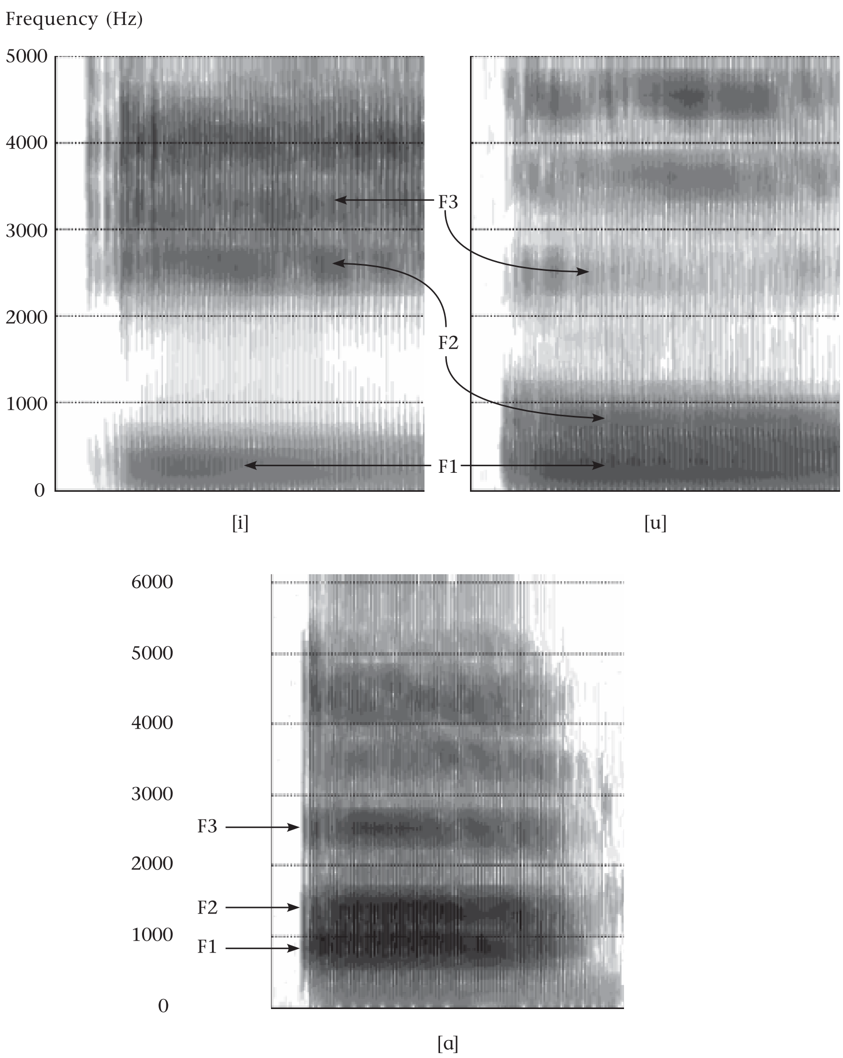
We can plot these vowels by the frequencies of their first two formants, as shown in (6). Note that if we put the origin (0,0) in the upper right-hand corner, the resulting diagram looks strikingly similar to the vowel chart in (1) in File 2.3. Thus we can see that the first formant corresponds inversely to the height dimension (high vowels have a low F1, and low vowels have a high F1), and the second formant corresponds to the advancement (front/back) dimension (front vowels have a high F2, and back vowels have a low F2).

A common method of visually representing acoustic properties of speech sounds is to use a **spectrogram.** Spectrograms are graphs that encode three acoustic dimensions: the vertical axis represents frequency, and the horizontal axis represents time. A third dimension is represented by degree of darkness, which indicates the amount of acoustic energy present at a certain time and at a certain frequency. Dark horizontal bands usually represent formants because formants are enhanced bands of energy at particular frequencies. In (7) we see spectrograms for the three vowels [i], [u], and [ɑ]. The arrows point out only the first three vowel formants, although there are more formants visible in these spectrograms. The horizontal lines in each of these displays mark off frequency in Hz by the 1000s. These spectrograms show visually the differences that we hear when we listen to these three vowels.

(6) Plot of the first formant (F1) against the second formant (F2) of some English vowels




(7) Spectrograms of the vowels [i], [u], [ɑ]



If you compare the spectrograms of [i] and [u] in (7), you will notice that while the first formant is very similar, the second formant is much higher for [i] than for [u]. This is because the second formant corresponds to tongue advancement; your tongue is much farther forward when pronouncing [i], so the corresponding formant is much higher. You can also see that the first formant of [i] and [u] is much lower than the first formant of [ɑ] because the former are high vowels while the latter is a low vowel (the first formant corresponds inversely to vowel height).

## 2.6.5 Stops

Spectrograms can clearly show other types of segments as well. In File 2.2, we described the articulatory properties of consonants in terms of their voicing and their manner and place of articulation. Stop consonants are produced by completely closing off the oral cavity with the lips or tongue, blocking the flow of air. This lack of airflow makes stops easy to detect on spectrograms because they are characterized by a lack of energy—hence a gap—in the display, as illustrated in (8). So, the acoustic characteristic of a stop (the silence we hear, or the blank space on the spectrogram) reflects its manner of articulation.

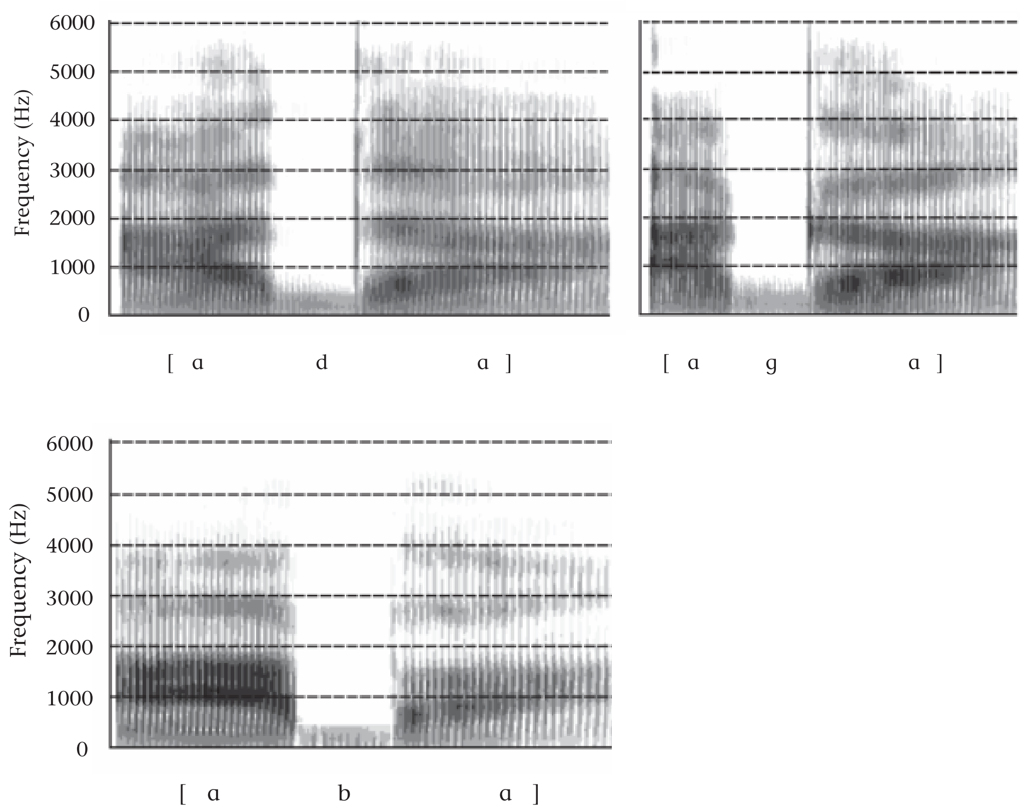
If a stop is voiced, the vocal folds will actually be vibrating during the closure, and some low-frequency noise is produced. This noise can be seen in (8) as the dark band at the very bottom of the spectrogram during the "silence" of the stop. This band is called the **voice bar.**

Voiceless stops never have this voice bar. In English, voiceless stops are also often further characterized by a period of **aspiration,** during which air rushes out of the mouth after the release of the stop closure and before the onset of the vowel. This aspiration, transcribed with a superscript [ʰ], as in [pʰ], can be clearly seen as the messy shading without clear formants that occurs between the gap of the consonant and where the formants of the vowel begin in the spectrogram of the word *pat* in (9).

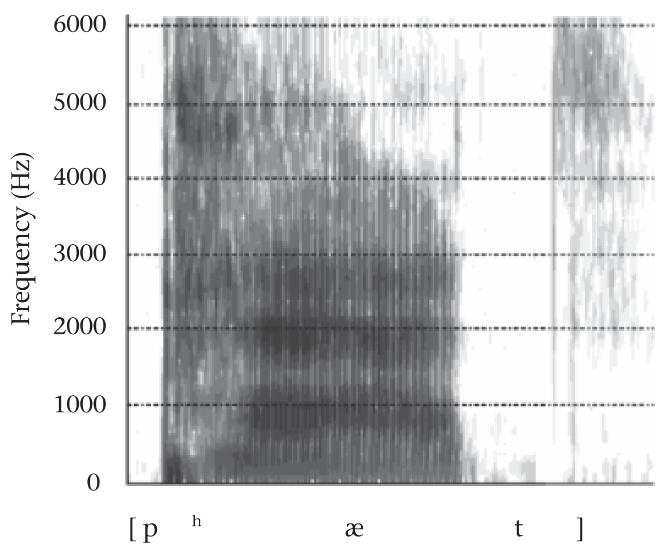
The acoustic information corresponding to place of articulation for a stop is found mostly in the vowels around it since, after all, the stop itself is essentially silence. When we pronounce a sequence like [ɑdɑ], the tongue can't move instantaneously from a low back tongue position to the alveolar ridge for the voiced alveolar stop and back to the vowel position. Rather, the tongue glides from one position to the next. Therefore, there are points in time when the tongue is in transition from the vowel to the consonant or the consonant to the vowel. Of course, this changing vocal tract shape affects the formants; as a result, during the early part of the second vowel the formants are also in transition toward their usual values. The spectrograms in (8) show vowel-stop-vowel sequences in which we can see moving formants reflecting the moving articulator. (The horizontal lines in each of these displays mark off frequency in Hz by the 1000s.)

We can determine the place of articulation of a stop by examining the frequency of the second formant at the juncture of the vowel and the consonant. For alveolar stops, the second formant of the vowel will be around 1700–1800 Hz going into or coming out of the consonant. Thus, in the spectrogram of [ɑdɑ] in (8), the second formant rises from the formant value of the vowel (the F2 of [ɑ] is about 1100 Hz) to about 1800 Hz just before the consonant [d] and falls back down afterward. For velar stops, the pattern will depend on what kind of vowel precedes or follows the consonant. For example, if the consonant is followed by a front vowel, the F2 will start high and then fall, but if the consonant is followed by a back vowel, the F2 will start fairly low, around 900 Hz or lower. In addition, just before the consonant, the second formant will rise and the third formant will lower, as if they would meet in the consonant. You can see this in the spectrogram of [ɑɡɑ] in (8). This is sometimes called a "velar pinch." For bilabial stops, F2 will be lower at the juncture between the consonant and the vowel than it is in the vowel itself. As the spectrogram of [ɑbɑ]

(8) Spectrograms of [ɑdɑ], [ɑgɑ], [ɑbɑ]




(9) Spectrogram of [pʰæt]



shows, the second formant lowers slightly just before the [b] and rises just afterward. These patterns allow us to distinguish the place of articulation of stops visually. The placement of the burst of air that follows the stop when the stop is released also gives information about the place of articulation of the stop.

## 2.6.6 Fricatives

Fricatives involve a new kind of sound that we have not dealt with up to this point. The difference between the noise found in vowels and in fricatives is that the sound in vowels has its source in the periodic vibration of the vocal folds (as discussed in Sections 2.6.2 and 2.6.3), while the sound in fricatives comes from the aperiodic, or random, turbulence of the air rushing through a small opening. Note in (10) that during the vowels there is a regular repetition (seen in the vertical stripes), while in the fricative portions there is no apparent pattern; it looks like static on a TV screen. In addition, note that this is the same kind of noise as that of the aspiration discussed in the previous section.

We find differences among English fricatives in the relative frequency of the noise (e.g., [s] has a higher frequency energy concentration in the frication noise than [ʃ]), in the amplitude (e.g., [s] is louder than [f]), which appears as darker shading on a spectrogram, and in the duration (e.g., [s] is longer than [z]). In Figure (10), you can see that the static-like coloring denoting the aperiodic energy of the fricative is centered between 6000 and 9000 Hz for [s], but is much lower, centered between 2000 and 4000 Hz, for [ʃ]. As with stops, the formant transitions from the consonant into the vowel are also used by listeners to determine the place of articulation.

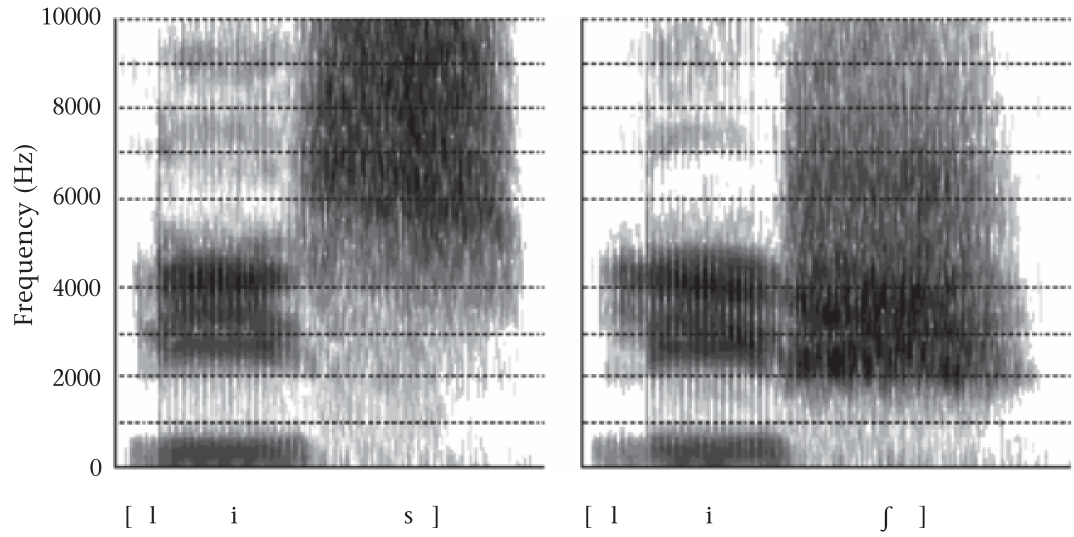
Voiced fricatives are interesting in that they combine periodic noise (the vocal folds are vibrating in a regular cycle) and aperiodic noise (there is turbulence from the air being forced through a small opening). Affricates are sequences of stop plus fricative both in their articulation and in their acoustic characteristics. A spectrogram of an affricate begins with a gap in the waveform, which is immediately followed by the aperiodicity of a fricative.

## 2.6.7 Nasals, Liquids, and Glides

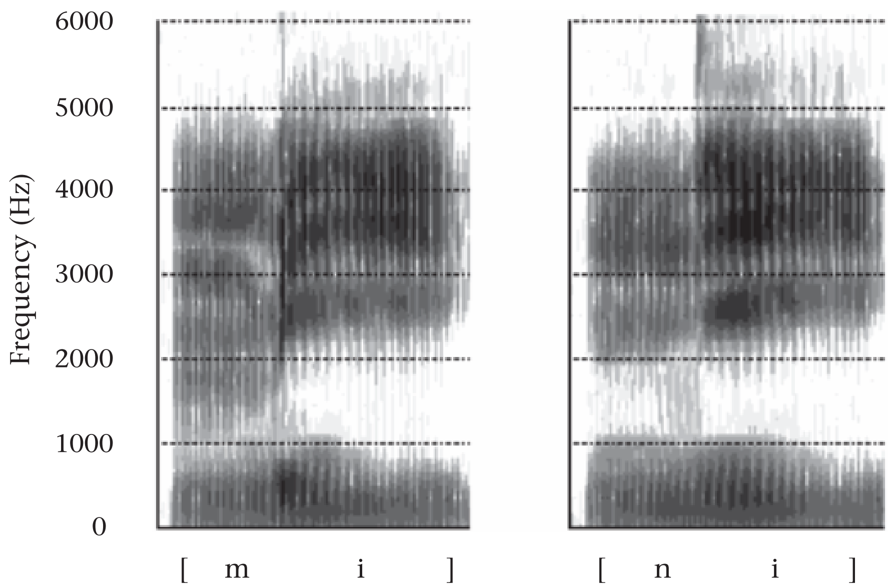
In the production of nasal consonants, the oral cavity is closed as if for a stop, but air escapes past the lowered velum through the nasal cavity. In acoustic terms, the nasal passage serves as the filter for the vocal source, just as the oral cavity acts as a filter in vowels. All nasal consonants have quite similar formants (see (11)), reflecting the shape of the nasal passage, which enhances some harmonics and damps others. Nasal formants are usually somewhere around 250, 2500, and 3250 Hz. The place of articulation of nasal consonants, however, is still cued by the transitions from the nasal into the vowel. Note that in (11), there is a lighter area (a lack of energy, caused by the damping of the nasal cavity) at around 1250 Hz for [mi] and around 1750 Hz for [ni].

Like nasals, liquids and glides have formants. Both nasals and liquids are characterized by an abrupt change in the spectrogram just before the consonant, which is very different from the gradual changes that mark the transition into stops and glides. The glide [w] has formants very similar to those of [u] (and [j] to [i]), but because consonants are shorter than vowels, the formants do not have time to reach those positions and stay there. Glides are sometimes appropriately called **semivowels**.

(10) Spectrograms of [lis] and [liʃ]



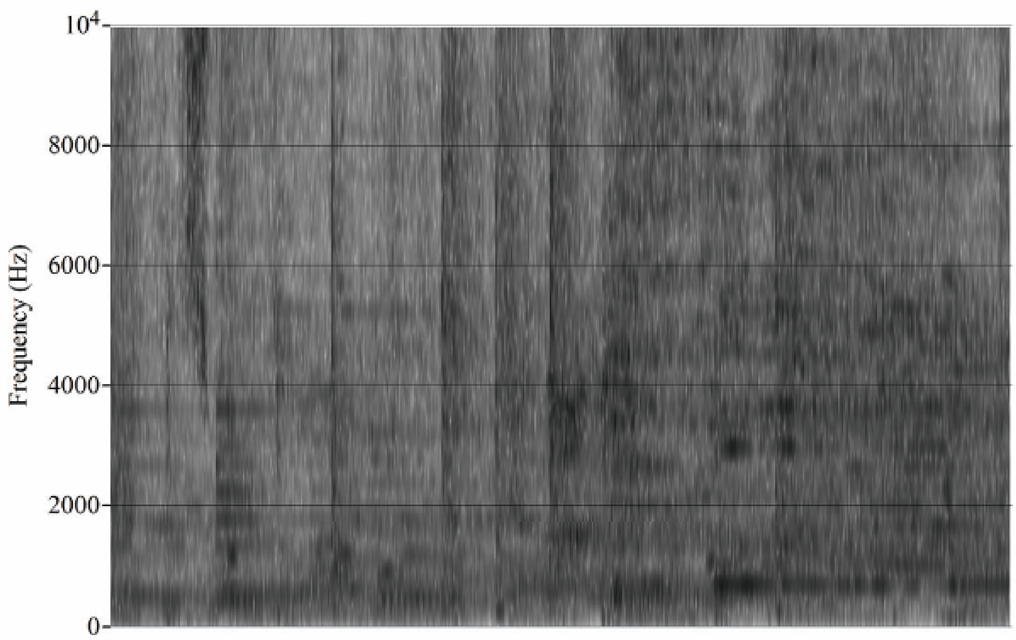

(11) Spectrograms of [mi] and [ni]



## 2.6.8 Interpreting Spectrograms

All of the sounds shown thus far in this chapter were recorded in a sound booth, spoken quite distinctly and almost in isolation, with few other sounds. But we do not usually speak this way and hardly ever listen to such speech. Most of the time, our speech is connected, and neighboring sounds can impact the cues of a sound a great deal (see Chapter 3). Sometimes our speech is rushed or mumbled, and conversations often take place with a great deal of background noise, making the sounds even harder to distinguish from one another. Therefore, while extremely useful for preserving, analyzing, and comparing speech, sound spectrograms of naturally occurring speech can be very difficult to interpret. For example, (12) shows the spectrogram of a recording made in the audience at a rock concert. Instead of the careful speech of one person, the 2.5-second spectrogram represents sounds produced by four singers, two guitars, a piano, drums, and thousands of screaming fans. It would be next to impossible to use this spectrogram for analysis of the singers' consonants and vowels, but fans in the audience would have little trouble understanding the lyrics and following along with the complex musical patterns made by the instruments while subconsciously filtering out the background noise of their cheering neighbors. This should highlight once again how incredibly talented our ears are at picking out speech sounds!

(12) Spectrogram from rock concert



**FILE 2.7**

# The Phonetics of Signed Languages

## 2.7.1 Extending *Phonetics* to Signed Languages

*Phonetics* was originally coined as a term used specifically to talk about the study of the sounds of language. However, phonetics has come to be the name of the subfield that deals with how language is produced, regardless of the modality of that production. Signs, which serve the same function as words in spoken languages, likewise have internal structure. Therefore, signs in any signed language are composed of discrete components, just like words in spoken language, and these components can be studied in the same way that vowels and consonants can be.

As has been the case for most of the preceding files of this chapter, the focus of this file will be on articulatory phonetics: how signs are produced. However, in the same way that phoneticians also study acoustic phonetics—the sounds themselves—and auditory phonetics—how sounds are perceived—linguists who are working on signed language phonetics may also take an interest in how signs are perceived or in the structure of the signs themselves, independent of how they are articulated.

## 2.7.2 The Parameters of Articulation in Signed Languages

The study of the phonetics of signed languages is relatively new. Thus, whereas linguists speak fairly confidently when they say, for example, that a significant feature in describing a consonant is place or manner of articulation, there is still some discussion about which attributes of a sign are significant. Nonetheless, there is a canonical set of parameters that are generally recognized in one way or another as being linguistically significant.

How do you know that a parameter is significant? Well, in evaluating spoken languages (if you speak one), the task is relatively easy. You know, for example, that *mitt* and *bit* are different words, and therefore the feature nasal must be important (because you are able to distinguish between [m] and [b], and they differ only in nasality). When we want to know whether some particular parameter is significant in a signed language, we can do much the same thing: we look to see whether a change to the articulation of that parameter can influence the identity of a sign. (This notion of how discrete yet meaningless units of language can come to affect meaning will be readdressed in Section 3.2.3.)

By performing this set of observations, we can conclude that there are five key **parameters** of articulation in signed languages: place of articulation, movement, handshape, hand orientation, and non-manual markers, each of which will be discussed in more detail below. The way that these features are organized, though, does not correspond directly to the way that features like nasal or rounded are organized. Rather, they themselves are segments. In the same way that a word will have some number of vowels and some number of consonants, a sign will have some number of movements and some number of places of articulation.

One fascinating difference between signed and spoken language is the manner in which their fundamental elements, called phones or **primes,** are combined into utterances.

In spoken languages, owing both to the nature of the speech mechanism and to the way that our brains process auditory input, phones are organized in linear temporal order; several phones cannot be produced at the same time. (Imagine trying to produce all the phones of a word at the same time! Furthermore, think about how difficult it is to understand three different people talking to you at the same time.) In contrast, a prime in ASL always occurs simultaneously with other primes. Primes cannot stand alone but must co-occur with primes from the other parameters. For example, one could not simply have a hand movement without also having the hand in a particular handshape or location. (It is possible not only to produce multiple primes at the same time, but also to interpret them. Imagine that you are shown a photograph, but that it flashes in front of you and then disappears immediately. You will be able to tell many things about the photograph, because our visual processing, unlike auditory processing, does allow us to clearly perceive multiple different things going on at the same time.) Because all five of the parameters of signing articulation discussed below are superimposed, they interact with one another in complex ways. One parameter may change while another stays the same, or two may change at the same time.

In the following sections we will describe each of the five parameters in more detail and provide several examples of each. Although the examples given in this file come from only one language, the same parameters are relevant for all signed languages.

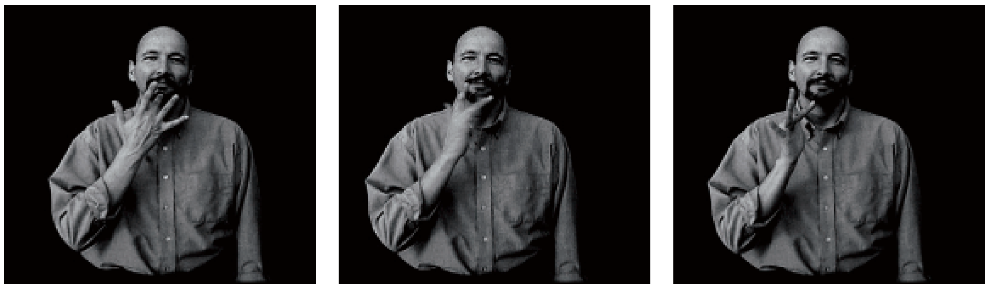
## 2.7.3 Location

The first parameter of sign articulation that we will consider is **location.** Clearly it is impossible to articulate a sign if the hands aren't somewhere! And we could imagine a system in which all gestures could be made anywhere at all and still have the same meaning (just as you can say a word at any volume at all and it still has the same meaning). How, then, do we know that location is important? We find pairs of words like the following. In the ASL signs for 'apple' (1a) and 'lucky' (2a), the location where the sign is made is at the chin. The sign for 'onion' in (1b) is the same in every way as the sign for 'apple' except that it is made near the eye. Similarly, the sign for 'clever' in (2b) is the same in every way as the sign for 'lucky' except that it is made starting at the forehead. Evidence like this tells us that location is significant.

(1) a. ASL: APPLE

b. ASL: ONION

(2) a. ASL: LUCKY



© 2006, William Vicars, www.Lifeprint.com. Adapted by permission.

b. ASL: CLEVER

© 2006, William Vicars, www.Lifeprint.com. Adapted by permission.

The examples in (1) and (2) have places of articulation that differ between the upper and lower halves of the face. Examples such as these are particularly clear in pictorial two-dimensional form; however, there are certainly other locations that contrast.

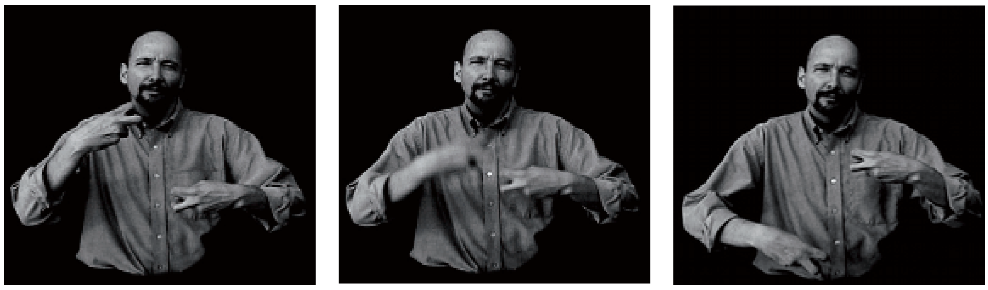
Every signed language has a particular "signing space": a general area in which signs may be produced. Obviously the outside range might be determined by how far away from your body you can stretch your arms, but most languages have a smaller space than this. For example, ASL has very few signs that are articulated below the waist. But place of articulation is a much more specific feature than just identifying a general area. A sign's place of articulation tells exactly where, relative to the signer's body, that sign must be articulated. Examples include [the front of the shoulder of the arm opposite from the hand making the sign], [the top of the bridge of the nose], [above the shoulder of the hand articulating the sign, but touching neither the shoulder nor the ear], and so on.

Interestingly, signing space can be expanded or reduced. If a signer is "whispering," he will reduce the signing space, bringing all places of articulation in closer to his center. This may also involve altering the location of some signs, articulating them in places closer in front of the torso than they normally would be. However, the places of articulation will still have the same sort of positions relative to each other. That is, in whispering, signs normally produced on the forehead will be lowered, while signs normally produced on the chin will also be lowered; every sign will come in toward the signer's center an equivalent amount. Similarly, if a signer is "yelling," he will increase his signing space and the amount of movement in his signs.

### 2.7.4 Movement

The second parameter is **movement.** The examples in (3) and (4) show two pairs of signs that are distinguished by the kind of movement they involve. In TOUGH, one hand begins higher than the other and moves rapidly downward until it is lower; PHYSICS is similar in many ways but involves the two hands moving toward each other.

(3) a. ASL: TOUGH (difficult)



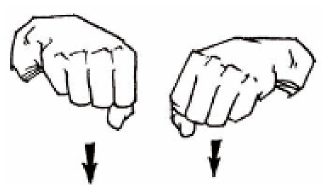


b. ASL: PHYSICS

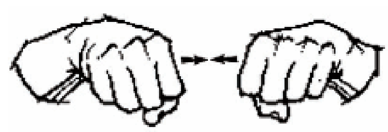


The signs CAN and SHOES likewise distinguish between vertical and horizontal movement (see (4)), though the vertical motion in CAN is different from the vertical motion in TOUGH. (Try to describe the difference in movement between these two signs.)

(4) a. ASL: CAN



b. ASL: SHOES





Some signs have movements that are designed to take a hand from one place of articulation to another. For example, the ASL sign for KING moves from the shoulder opposite the signing hand to the top of the hip on the same side as the signing hand. This is different from, for example, the sign for TOUGH above, because in TOUGH what matters is the type of movement itself, more than the precise starting and stopping location.

A third type of movement has to do with ways that the wrist or fingers move and does not actually require any change in place at all. For example, in the ASL sign for YES, the wrist moves up and down (as though it were a head nodding), and in the ASL sign for WAIT, the fingers waggle back and forth, but the hands do not move. Other such movements may include finger circling or one hand tapping another body part.

One interesting thing about movement is that it functions a little bit like vowels in spoken language. You can often understand a word or sentence (in its written form) without vowels; similarly, a signer can often understand a sign or sentence without movement. Nonetheless, just like vowels in spoken languages, movement is a critical part of articulation in signed languages.

### 2.7.5 Handshape

Third, we will look at **handshape.** In (5) you see four signs of American Sign Language, each with no movement and with the same place of articulation (touching the chin). What differs is the shape of the hand: which fingers are extended, whether the fingers are bent or straight, the position of the thumb, whether fingers are touching, and so on. In (5), the four different handshapes give four different meanings to the four signs that they are a part of.

(5) Examples of signs in ASL differing only in handshape

a. COLOR

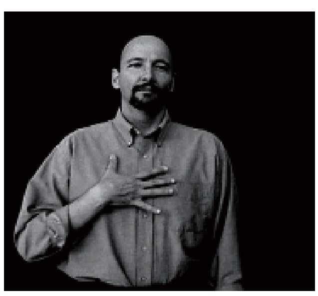
b. ORANGE

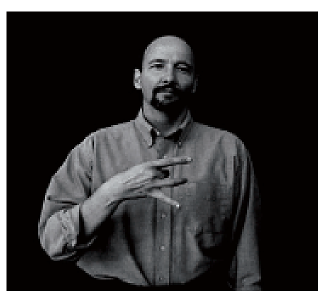
c. MISS

d. WRONG

In order to see one way that handshape can interact with movement, consider the two ASL signs in (6). Here, although both LIKE and WHITE begin with the same handshape, they end with different handshapes, because the handshape changes during the movement.

(6) a. ASL: LIKE





© 2006, William Vicars, www.Lifeprint.com. Adapted by permission.

b. ASL: WHITE

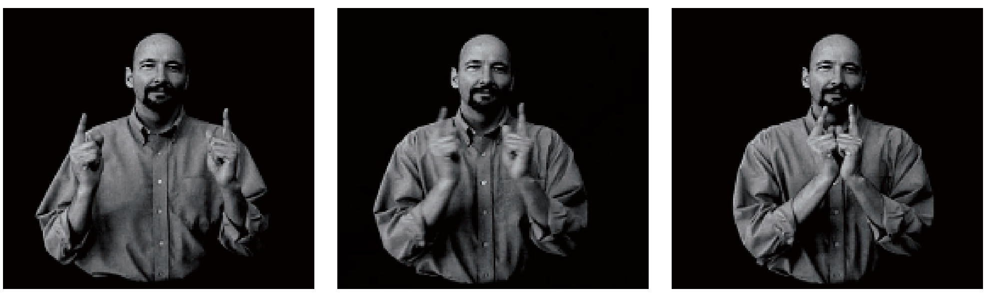


The two signs in (6) also serve to make the point that one sign can contain more than one handshape.

## 2.7.6 Orientation

The fourth parameter that has to do with the way that the hands are used is **orientation**: the direction that the palm of the hand is facing. In both (7a) and (7b), the hands are facing toward each other; however, in (7a) the two hands are pointing left and right, whereas in (7b) they are facing toward the speaker and away from the speaker.

(7) a. ASL: MEET (the uninflected verb)





b. ASL: I MEET YOU



Of course, even in two-handed signs, the hands need not face each other; in the signs for CAN and SHOES in (4), the palm orientation is [facing down]. In the signs for LUCKY and CLEVER in (2), there is a change of orientation during the sign: these two signs begin with the palm facing the speaker and end with the palm facing away from the speaker.

### 2.7.7 Non-Manual Markers

The fifth and final parameter of signed language is the use of non-manual markers. Non-manual markers include any gestures, such as facial expressions or head movements, that are not made with the hands. The examples in (8) show the ASL signs for LATE and NOT YET. While the location, movements, shape, and orientation of the hands are the same in the two signs, NOT YET also includes non-manual markers: a slight negative shake of the head and the tongue sticking out slightly.

(8) a. ASL: LATE

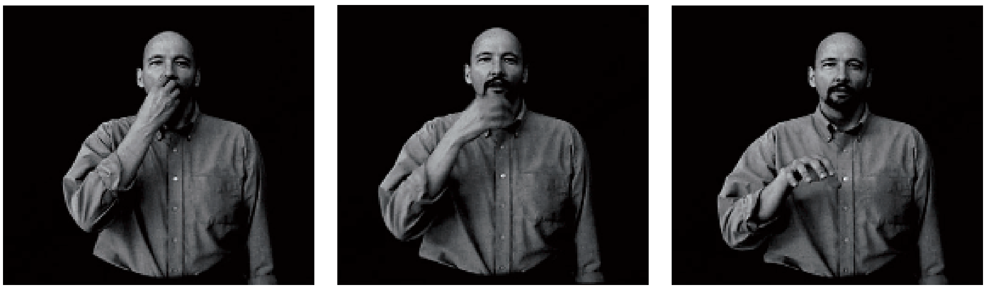
b. ASL: NOT YET

Another example of this parameter is the use of pursed lips in the ASL sign REALLY-SKINNY. The sign in question also has a manual component: two hands, each with the pinkies extended, begin with the pinkies touching and then move away from each other. However, if just this manual part is performed, without the lips being pursed, then the entire word hasn't been articulated. It would be like leaving a segment out of some spoken word: saying [fut] instead of [flut] for *flute,* for example. If somebody were to say, "I play the [fut] in the school orchestra," you would know what they meant, but you would also know that they hadn't articulated the word properly. A second example would be not producing a tone change in a word that required one in a language like Mandarin Chinese, in which tones are components of word production.

In other cases, non-manual markers act as more of a suprasegmental feature like intonation. For example, there is a particular intonation that we associate with questions like *Where do you live?* In ASL there is also a suprasegmental feature that indicates such a question: it includes inclining the head forward and lowering the eyebrows.

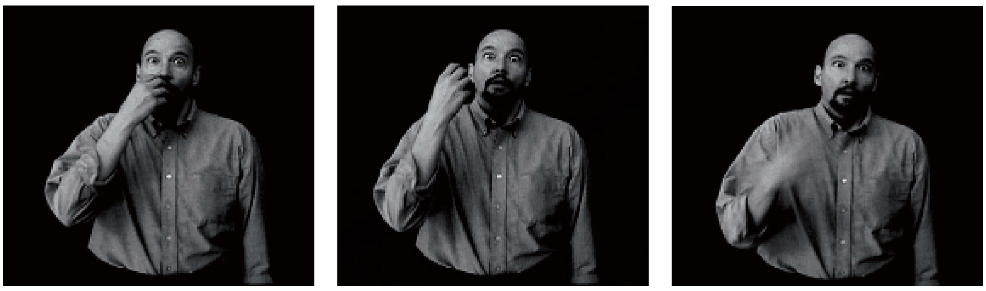
Non-manual markers can be used to modify signs in other ways as well; for example, compare the signs for HOT in (9a) and VERY HOT in (9b). Notice how the signer's face is different when he articulates VERY HOT from when he articulates HOT. (There are other differences in the production of these two signs as well, but for now just pay attention to the signer's facial features.)

(9) a. HOT





b. VERY HOT





Sometimes, the non-manual marker/facial expression is the only way a sign is indicated. For instance, just the non-manual marker of NOT YET (the tongue sticking out a bit and a slight shake of the head) can carry the full meaning of NOT YET.

## 2.7.8 Phonetic Inventories in Signed Languages

In File 2.4, it became clear that different languages make use of different inventories of sounds. Some languages have front rounded vowels or consonants with a uvular place of articulation, but English has neither; English is rather unusual in having a retroflex liquid; and so on. All spoken languages have some kinds of consonants and vowels, but the sets of consonants and vowels differ from language to language. The same is true of signed languages. Every language has handshapes, kinds of movements, places of articulation, orientations, and non-manual markers, but not every one is available in every sign language.

For example, in Taiwan Sign Language (TSL), there is a handshape that is called the "dragon" handshape, formed by sticking the pinky and index finger up while bending the middle finger and ring finger in to meet the thumb. If you try to make this handshape, you will find that it is not terribly difficult to produce. Nonetheless, this handshape is not available in the inventory of handshapes that are used in ASL. A second example is the handshape formed by making a fist and extending your ring finger: TSL makes use of it, but ASL does not. Conversely, the ASL "T" handshape, which is formed by making a fist and sticking the thumb between the index finger and middle finger (as though you were playing "I got your nose" with a young child), is a handshape that is not available in TSL. There are other handshapes that appear in neither ASL nor TSL but that do occur in other sign languages. A more profound difference is that in TSL, the elbow can be an active articulator, whereas in ASL the forearm and elbow can only be used as passive articulators. (To conceptualize what this means, think about your mouth: your tongue is an active articulator because it moves, but your alveolar ridge is a passive articulator because it is involved in articulation only when your tongue touches it.)

The same sort of thing (primes that are available in one signed language but not another) occurs for kinds of movement and places of articulation. Some languages have a movement that is a side-to-side twisting of the wrist; others do not. Some sign languages have [crown of the head] as a place of articulation; others do not.

There are many things, of course, that you can do with your hands and arms—just as there are many things you can do with your mouth. Some of these, such as swallowing, whistling, throwing a ball, or brushing at a mosquitos, are nonlinguistic, while others may be linguistic. It is important to remember, though, that just because a certain kind of articulatory gesture may have linguistic applications does not mean that any given language necessarily uses it.

### 2.7.9 Studying and Analyzing the Phonetics of Signed Languages

In the previous files, a number of innovations have been described that help researchers to discuss, describe, and research the articulation of spoken languages. There have been fewer technological innovations for the study of phonetics in signed languages, in part because the sign language articulators are large, slow, and not covered by your cheeks. In other words, they are a lot easier to study in a straightforward way than are the articulators of spoken languages!

Another reason, though, is that, as we mentioned above, the study of the phonetics of signed languages is simply newer than the study of spoken language phonetics. Of course, one tool that has been very helpful is simple video recording, which allows researchers to look at the same segments over and over again. More sophisticated technology involves attaching sensors to various parts of signers' hands, arms, face, and so on. The sensors' positions and movements can then be recorded and sent to a computer to allow precise measuring of, for example, amount of movement, precise tilt and orientation, exact distance between hands and between hands and the body, and so on. Of course, as this field of study continues to grow, more instruments and tools are certain to follow.