3

The Sounds of Languages: Phonetics

What Do You Think?

- A whiz at reading, your niece Nina reports one day that English has five vowels—*a, e, i, o,* and *u.* You think it must have more than five vowel sounds. How would you figure out just how many vowel sounds English actually has?
- Roommate Ron wants to know whether English has 26 sounds to match the 26 letters of the alphabet. You figure English must have more than 26 sounds and point out that it lacks a letter to represent the initial sound in *thin* so is forced to use two letters. Ron asks for other examples where two letters represent a single sound because there isn't an English letter for the sound. He doesn't want examples like *physics,* where an <f> would do, or *pneumonia,* where an <n> would do. What examples can you provide?
- Citing *put* and *putt* as a pair of English words that are pronounced differently and spelled differently but whose spelling difference doesn't correspond to the pronunciation difference, your friend Fred claims English has many similar pairs and challenges you to name just one other one. Can you do it?
- Neuroscience major Nicole claims that the organs that produce human speech in the vocal tract evolved so that humans could talk. Bio major Bill disagrees. He says the organs that produce human speech evolved for a different purpose altogether and that human speech merely takes advantage of those organs for talking. Who's right?

Sounds and Spellings: Not the Same Thing

As a reader of English, you are accustomed to seeing language written down as a series of words set off by spaces, with each word consisting of a sequence of separate letters that are also separated by spaces. You readily recognize that words exist as separate entities made up of a relatively small number of discrete sounds. The words *sat* and *put*, for example, have three sounds each, while *until* has five and *untold* has six. Somewhat less obvious is the number of sounds in *speakers, series, letters,* and *sequence,* which do not have the same number of letters and sounds. This lack of correspondence is common in English. *Cough* has three sounds but is spelled with five letters; *freight* uses seven letters to represent only four sounds.

Through with seven letters and *thru* with four are alternative spellings for a word with three sounds. *Phone* and *laugh* have three sounds each, represented by five letters. *Delicacy*, with an equal number of sounds and letters, uses the letter <c> to represent two sounds—the first a *k*-like sound, the second an *s*-like sound.

Because of the close association between writing and speaking in the minds of literate people, it is important to stress that in this chapter we are interested in the *sounds* of spoken language, not in the letters of the alphabet or other ways that represent those sounds in writing.

Same Spelling, Different Pronunciations

Observe the variety of pronunciations represented by the same letter or series of letters in different words. Consider the pronunciations of the following English words, all of which are represented in part by the letters <ough>:

cough	"koff"
tough	"tuff"
bough	"bow"
through	"thru"
though	"tho"
thoroughfare	"thurrafare"

Though the precise sounds in these words may vary somewhat among English speakers, still the lesson of the distant relationship between sounds and letters is clear. The <ough> spelling represents at least six pronunciations in English, as indicated in Figure 3.1.

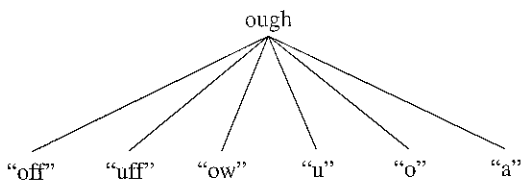


Figure 3.1 Same Spelling, Different Sounds

Same Pronunciation, Different Spellings

Other sets of English words are pronounced alike but spelled differently, as schoolchildren learn when they are taught sets of homophones (or homonyms) such as *there/their, bear/bare, led/lead, core/corps,* and *to/two/too.*

Consider the words in Figure 3.2, where nine different spellings represent the sounds of the word *see* (and *situ* and *cee*—the name of the third letter of the alphabet—could be added to the list). Notice that the letter <x>, as in *sexy* and *foxy*, stands for the two sounds [k] and [s] as represented in *folksy*.

Compare the sound and spelling of *woman* and *women* and you'll note that the difference in the letters <a> and <e> does not represent a difference in pronunciation because the second syllables <-man> and <-men> are pronounced alike in these words. On the other hand, the letter <o>, which does not change, represents two sounds (like the <oo> of *wood* in *woman* and like the <i> of *win* in *women*). The pair *Satan* (the devil) and *satin* (the cloth) illustrates the same point: the <a> of the first syllable represents two sounds, but the <a> and <i> spellings of the second syllable represent the same sound. The same point can be made with *loose* and *lose*, where the only pronunciation difference is in the final sound ([s] versus [z]), while the only spelling difference is in the similarly pronounced vowels.

The playwright George Bernard Shaw was a keen advocate of spelling reform and highlighted the problems in establishing correspondences between English sounds and spelling when, as has been claimed, he provocatively alleged that *fish* could be spelled <ghoti>: the <gh> as in *cough*, the <o> as in *women*, and the <ti> as in *nation*. Despite the efforts of Shaw and other reformers, English spelling has remained basically unchanged. You can see slight success at simplification in such isolated spellings as *thru*, *nite*, and *foto*, though not even these examples have been widely adopted for the more traditional *through*, *night*, and *photo*.

Whys and Wherefores of Sound/Spelling Discrepancies

Here are five reasons for the discrepancy between pronunciations and written representations for many English words.

1. Written English has diverse origins with different spelling conventions:
   - *Anglo-Saxon* The system that evolved in Anglo-Saxon England before the Norman Invasion of 1066 gave us such spellings as *ee* for the sound in words like *deed* and *seen*.
   - *Norman French* The system that was overlaid on the Old English system by the Normans, with their French writing customs, gave us such spellings as *queen* (for the earlier Old English *cwene*) and *thief* (for the earlier *theef*).

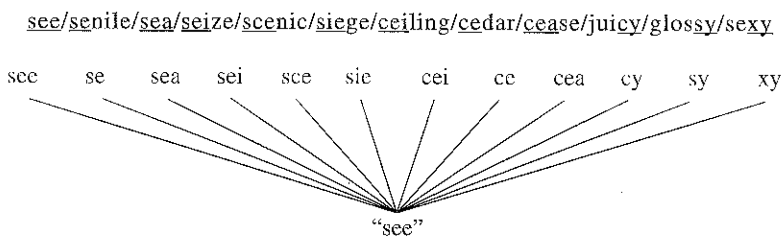


Figure 3.2 Different Spelling, Same Sounds

- *Dutch* Caxton, the first English printer, who was born in England but lived in Holland for 30 years, gave us spellings influenced by Dutch, such as *ghost* (which replaced *gost*) and *ghastly* (which replaced *gastlic*).
- *Spelling reform* During the Renaissance, attempts to reform spelling along etymological (that is, historically earlier) lines gave us *debt* for the earlier *det* or *dette* and *salmon* for the earlier *samon.*

2. A spelling system established several hundred years ago is still used to represent a language that continued—and continues—to change its spoken form. For example, the initial <k> in words like *knock, knot, know,* and *knee* was earlier pronounced, and so was the <gh> in *night* and *thought.* As to vowels, pronunciation changes have led to discrepancies like those represented in *beat* versus *great* and *food* versus *foot.*
3. English is spoken differently around the world (and in different regions of a given nation), despite relatively uniform standards for spellings. Such spelling uniformity facilitates international communication, but it also increases the disparity between the way the language is written and spoken.
4. Word parts (morphemes) may be pronounced differently depending on their adjacent sounds and stress patterns. In *electric,* the final <c> represents [k] as in *kiss,* but in *electricity* that same <c> represents [s] as in *sit.* In *senile,* <i> is pronounced as in *island,* but in *senility* as in *ill.*
5. Spoken forms may differ across social situations. The writing system incorporates some degree of variation (*do not* versus *don't* and *it was* versus *'twas*), but there is less tolerance for spellings like *gonna* ('going to'), *wanna* ('want to'), and *gotcha* ('got you') and still less for *j'eat* ('did you eat?') and *woncha* ('won't you?'). Variable spellings for the same expression would force readers to determine the idiosyncratic key of the represented speech before arriving at meaning instead of relying directly on a familiar spelling, as adult readers normally do.

Advantages of Fixed Spellings Some disadvantages of an inconsistent set of sound-spelling correspondences are obvious, but the advantages are also substantial. Consider Chinese, in which many written characters make little or no reference to sounds but directly symbolize meanings—much as numerals like 3 and 7 and symbols like + and % do in European languages. Using such characters, groups of people whose spoken languages are mutually unintelligible can nevertheless communicate well in writing, as is the case between speakers of Cantonese and Mandarin Chinese. As a parallel, consider that the symbol 7 has a uniform meaning across European languages, even though the word for the concept is pronounced and spelled differently: *seven* in English, *sept* in French, *sette* in Italian, *sieben* in German, and so on. Similarly, the fact that English spelling is somewhat independent of pronunciation is not a bad thing when you consider that English has strikingly different dialects from Newfoundland to New Zealand and from Jamaica to Johannesburg. Despite diverse pronunciations around the globe, a uniform written word is associated with a single set of meanings. Moreover, in a language with different pronunciations for the same element of meaning, stable spellings can contribute to reading comprehensibility—as in *musical/musician, electrical/electricity,* and even the <s> of *cats* and *dogs* (pronounced [s] and [z], respectively).

Independence of Script and Speech It's important to distinguish between the sounds of a language and the way they are represented in writing. The independence of sounds and spellings is highlighted by remembering that a single language may be represented by more than one writing system. The language sometimes known as Hindi-Urdu is written by Hindus living in India in Devanāgarī, an Indic script that derives from Sanskrit, but in Arabic script by Muslims living in Pakistan and parts of India. Compare the Hindi चाय and Urdu چائے spellings for the same word *chai*, meaning 'tea.' Sometimes, too, people adopt a new writing system for their language. Early in the twentieth century, the government of Turkey changed the orthography (the technical term for a writing system) for representing Turkish from an Arabic script to one based on the Roman alphabet.

Try It Yourself Examine the "Polling Place" photograph in Chapter 1 (page 4). The left-hand list represents Chinese, Japanese, and Korean, and the Japanese word for polling place happens to use only kanji. What do you notice about the first two characters of the Chinese and Japanese versions?

Sometimes languages use different scripts for different purposes. Imagine sending an international telegram in a language that uses a script other than the Roman alphabet—Japanese, Korean, Greek, Russian, Persian, Thai, or Arabic, for example. Rather than using their customary orthographies, speakers of these languages use the Roman alphabet to send telegrams internationally. Even within a country, an alternative writing system may be needed. In China, each character has a four-digit numeral assigned to it and these numerals are sent telegraphically and then "translated" back into Chinese characters. Sometimes a language uses more than one writing system for different aspects of writing. Japanese draws upon three kinds of writing: *kanji*, based on the Chinese character system, in which a symbol represents a word independent of its pronunciation, and two syllabaries. A *syllabary* is a writing system in which each symbol represents a spoken syllable. Throughout the world there are discrepancies between sounds as spoken and as represented orthographically.

© Nan-hsin Du

Bilingual Sign, Xinjiang Autonomous Region, China. Written alternately in Arabic script (for Uyghur) and Chinese characters (for Chinese), with Arabic numerals in both. Neither language is related to Arabic.

In this chapter, we focus on the human vocal apparatus and the sounds it produces for speech. In Chapter 4 we'll examine the nature of sound systems in human language.

Phonetics: The Study of Sounds

Phonetics is the study of the sounds made in the production of human speech. It has two principal branches.

- *Articulatory phonetics* focuses on the human vocal apparatus and describes sounds in terms of their articulation in the vocal tract; it has been central to the discipline of linguistics.
- *Acoustic phonetics* uses the tools of physics to study the nature of sound waves produced in human speech. Acoustic phonetics is increasingly important in developing machines and computer software for speech recognition, as well as in voice identification and voice-initiated mechanical operations.

Our discussion in this chapter will be limited almost exclusively to articulatory phonetics—the nature of human sounds as they are produced by the vocal apparatus. But we'll make occasional mention of acoustic phonetics and in the references at the end of the chapter we'll identify useful sources for exploring modern technologies.

The late Professor Peter Ladefoged, second from right, coaches Rex Harrison, at left, on the set of the 1964 film *My Fair Lady*. Ladefoged later became president of the International Phonetic Association and of the Linguistic Society of America. The phonetic symbols on the chart were invented by Alexander Melville Bell (father of Alexander Graham Bell) and published in 1867 in his book *Visible Speech*. They preceded the International Phonetic Alphabet.

Phonetic Alphabets

To refer to the sounds of human language in terms of their articulation, phoneticians have developed descriptive techniques that avoid the difficulties of characterizing sounds across languages in terms of ordinary writing systems. You already know it is impossible to use customary written representations to analyze sound structure because, even within a single language, some sounds correspond to more than one letter and some letters to more than one sound. Besides that, the same letter or written character can represent different sounds in different languages. So an independent system for representing the sounds of human languages is a scientific necessity.

In scientific discussion, the requisite characteristics of symbols for representing sounds are clarity and consistency. The best tool is a phonetic alphabet, and the most widely used one is the International Phonetic Alphabet (IPA). The IPA aims to provide a unique written representation of every sound in every language, and it has been estimated that there are about 800 different consonant sounds and a couple hundred vowel sounds in the languages of the world.

A list of symbols used to represent the consonant sounds of English is given in Table 3.1. It provides the phonetic symbol for each sound and underscores the relevant parts of the words exemplifying those sounds. Taking advantage of an option allowed by the International Phonetic Association (the developer and guardian of the IPA), we use an ordinary printed <r> in this book to represent the initial sound of *ride* and the second sound of *pride*, but you should note that the association assigns the symbol [r] to a different sound. (We've indicated the standard IPA symbol for the initial sound of *ride* in parentheses.) The words illustrate word-initial, word-medial, and word-final occurrences of the sounds.

Table 3.1 *English Consonants Arranged by Position in Word*

Phonetic Symbol	Initial	Medial	Final
p	pick	caper	tap
b	bit	labor	tab
t	tick	meter	bat
d	dish	medal	pad
k	kiss	sicker	lick
g	geek	dagger	bag
f	fit	beefy	chief
v	vim	saving	grave
θ	thin	author	breath
ð	then	leather	breathe
s	sit	mason	kiss
z	zest	posit	buzz
ʃ	shed	rashes	rush
ʒ	genre	measure	rouge
ʧ	chip	kitchen	peach
ʤ	jet	bludgeon	fudge
m	mop	dummy	broom
n	nip	sunny	spoon
ŋ	______	singer	sing
h	hit	ahoy	______
j	yes	beyond	toy
r (IPA ɹ)	rest	berry	deer
l	last	silly	mill
w	wish	away	______

The Vocal Tract

The processes used by the vocal tract in creating a multitude of sounds are akin to those of wind instruments and organ pipes, which produce different musical sounds by varying the shape, size, and acoustic character of the cavities through which air passes once it leaves its source. Every speech sound you make differs from every other one because of a unique combination of features in the way you shape your mouth and tongue and move them and other parts of the vocal apparatus while making it. We will look at the parts of the vocal tract and show how they work together to produce sounds. Examine the simple drawing of the vocal tract in Figure 3.3.

How are speech sounds made? First, air coming from the lungs passes through the vocal tract, which shapes it into different speech sounds. The air then exits the vocal tract through the mouth or nose or both.

Despite the fact that speakers of all languages have the same vocal apparatus, no language takes advantage of all the possibilities for forming different sounds, and there are striking differences in the sounds that occur in different languages. For example, Japanese and Thai lack the [v] sound of English *van*, and Japanese lacks the [f] sound of *fan*. Thai lacks the sounds represented by <g> in *gill*, <z> in *zebra*, <sh> in *shell*, <s> in *measure*, and <j> and <dg> in *judge*. French, Japanese, and Thai lack the different sounds represented by <th> in *ether* and *either*.

Just as some languages lack sounds that English has, other languages have sounds that English lacks, and some languages that share a particular sound may show a different distribution of it within words. You know that English lacks the trilled *r* of Spanish and Italian and that German has a sound at the end of words like *Bach* 'stream' and *hoch* 'high' that does not occur in the inventory of English sounds. Arabic has a sound similar to the German <ch> of *Bach*, but in Arabic it can occur word initially. A similar (but not identical)

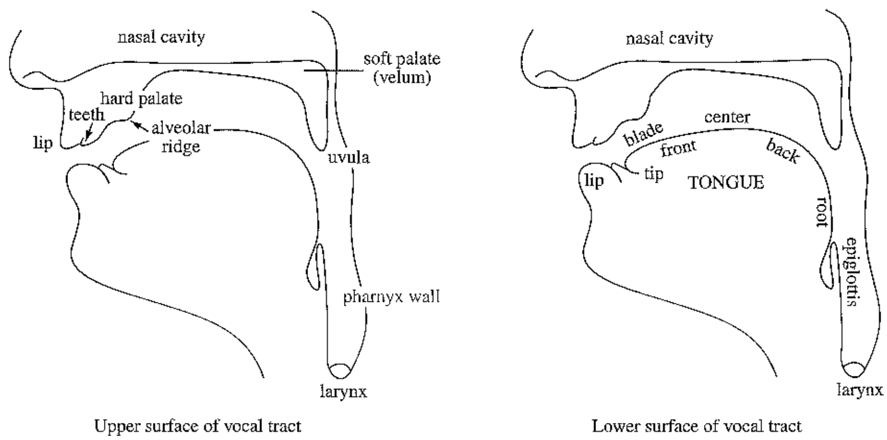


Figure 3.3 The Vocal Tract

Source: From LADEFOGED. *Course in Phonetics*, 4E. © 2001 Cengage Learning.

sound occurring word finally in the German word *ich* occurs in English (for those dialects that pronounce the<h>) in the initial sound of *human* and *huge*. Still, it can be tough for English speakers learning German to pronounce the sound in a word like *ich* because English doesn't permit that sound at the end of a word.

As you explore the inventory of sounds in the sections below, don't be shy about using your vocal tract to produce the sounds described. It's important to pronounce them aloud, noting the shape of your mouth and the position of your tongue for each sound. That experience will familiarize you with the reference points of phonetics, make the discussion much easier to follow, and enhance your confidence as you master articulatory phonetics.

We will continue to use square brackets to enclose the symbols that represent sounds. Thus [t] will symbolize the initial and final sounds in *tot*, [d] the initial and final sounds in *did*, and [z] the initial sound in *zebra*, the medial consonant in *busy*, and the final sound of *buzz* and *dogs*.

Describing Consonant Sounds

A **consonant** is a speech sound produced by a partial or complete closure of part of the vocal tract, thus obstructing the airflow and creating audible friction. Speech sounds can be characterized in terms of their *articulatory* properties—by *where* in the mouth and *how* they are produced, and consonants can be described in terms of three properties:

- **Voicing** (whether the vocal cords are vibrating or not)
- **Place of articulation** (where the airstream is most obstructed)
- **Manner of articulation** (how the airstream is obstructed)

Voicing

Begin by pronouncing a long, continuous [zzzzz] sound and alternating it with a long, continuous [sssss] sound. You'll notice that the position of your tongue within your mouth remains exactly the same, even though the sounds are noticeably different. You can feel the difference by touching your throat at the **larynx** (voice box) while saying [zzzzz sssss zzzzz sssss]. The vibration you feel from your larynx when you utter [zzzzz] but not [sssss] is called **voicing**; it is the result of air being forced through a narrow aperture (called the **glottis**) between two mucosal folds (the vocal cords) in the larynx. When the vocal cords are held together, the air forced through them from the lungs causes them to vibrate. It is this vibration, called phonation or *voicing*, that distinguishes [z] from [s] and enables speakers to differentiate between these two otherwise identical sounds. Using these highly similar but distinct sounds enables speakers to create words that differ by only a single feature of voicing on a single sound but carry quite different meanings, as in *peas* and *peace*, *fleas* and *fleece*, *zip* and *sip*, and *buzzes* and *buses*.

The Vocal Cords and Voicing

Human beings have no organs that are used *only* for speech. The organs that produce speech sounds have evolved principally to serve the life-sustaining processes of breathing and eating. Speech is a secondary function of the human "vocal apparatus" and is sometimes said to be parasitic on these organs. The vocal cords exemplify the parasitic nature of the vocal apparatus for speech: the primary function of the two folds of the "vocal cords" is to keep food from going down the wrong tube into the lungs.

With respect to speech, vibration of the vocal cords is what distinguishes voiced and voiceless sounds. You can perceive the difference between voiced and voiceless consonants by alternating between the pronunciations of [s] and [z] or [f] and [v] while holding your hands clapped over your ears. You should be able to tell from pronouncing the words *thin* and *thirty* (which begin with [θ]) and *then* and *those* (which begin with [ð]) which of the two sounds—[θ] or [ð]—is voiced. Check your conclusions against Table 3.6.

Other pairs of sounds are likewise characterized by a voiced versus voiceless contrast. Consider [f] and [v], as in *fine* and *vine* or *proof* and *prove*: both sounds are produced by forcing air through a narrow aperture between the upper teeth and the lower lip. The difference is that [f] is voiceless and [v] is voiced. Other voiceless/voiced pairs include [p] and [b] as in *pet* and *bet* or *rope* and *robe* and [t] and [d] as in *ten* and *den* or *net* and *Ned*.

Manner of Articulation

Besides having a voicing feature, consonants can be characterized as to their **manner of articulation**. In pronouncing [s] and [z], air is *continuously* forced through a narrow opening at a place behind the upper teeth. Compare the pronunciation of [s] and [z] with the sounds [t] and [d]. Unlike [s] and [z], [t] and [d] are not pronounced by a continuous stream of air passing through the mouth. Instead, the airstream is completely stopped behind and above the upper teeth and then released in a small burst of air. For this reason, [t] and [d] are called *stops*, and because the air is released through the mouth (not the nose), they may be called *oral stops*. Sounds like [s] and [z] that are made by a continuous stream of air passing through a narrowed passage in the vocal tract are called *fricatives*.

Try It Yourself Pronounce the sounds [p], [b], [f], [v], [θ], and [ð] to determine which are stops and which are fricatives.

Place of Articulation

Of the sounds analyzed so far, [s] and [t] are voiceless, [z] and [d] voiced. All four are pronounced with the point of greatest closure immediately behind and just above the upper teeth. Pronounce *ten* and *den* aloud, feeling where the tip of your tongue touches the top of your mouth for the initial and final consonants. Both words start and finish at the alveolar ridge. Because [t], [d], and [n] are articulated at the alveolar ridge, they are called **alveolars**. Also articulated at the alveolar ridge are [s] and [z], as you can notice by pronouncing the words *bus* and *buzz*. ([s] and [z] are fricatives, as you should be able readily to tell, whereas [t] and [d] are stops.)

There are three major **places of articulation** for English stops: the lips, the alveolar ridge, and the soft palate (or velum). If you say *pin* and *bin*, you'll notice that for the initial sound in each word air is built up behind the lips and then released. Thus the point of greatest closure is at the lips, and for that reason [p] and [b] are called **bilabial** stops (*bilabial* means 'two lips').

Try It Yourself Compare your pronunciation of [p] and [t]. Both are voiceless, so what is the difference between them? Pronounce word pairs like *pin* and *tin* or *tripe* and *trite* as other examples.

Try It Yourself Identify the pairs of stops pronounced at the lips, the alveolar ridge, and the velum.

Attend to the pronunciation of the first sound of *kin*, and you'll notice that [k], like [p] in *pill* and [t] in *till*, is a voiceless stop, but it differs from [p] and [t] in its place of articulation: [k] is a **velar** because it is pronounced with the back of the tongue touching the roof of the mouth at the velum (the soft palate); [k] is a voiceless velar stop.

Corresponding to the voiceless stops [p], [t], and [k] are three voiced stops: [b] as in *bib* is a voiced bilabial stop; [d] as in *dude* is a voiced alveolar stop; and [g] as in *gig* is a voiced velar stop. English has three pairs of stops, each pair pronounced at a particular place of articulation, with one member of the pair voiced and the other voiceless.

Besides lips, alveolar ridge, and velum, English takes advantage of other articulators to produce consonant sounds. The <th> of *thin* is a fricative pronounced with the tongue between the teeth. It is a voiceless **interdental** fricative, represented by the Greek letter theta [θ]. [ʃ] (the sound represented by <sh> in *shoot* and *wish*) and [ʒ] (the final sound in *beige*, the middle consonant in *measure*, and the initial sound in a tiny number of borrowed words like *genre*) are pronounced between the alveolar ridge and the velum (or palate) and are called **postalveolars**. [ʃ] is a voiceless postalveolar fricative, [ʒ] a voiced postalveolar fricative.

Kinds of Consonant Sounds

To repeat what was said above, consonants are produced by partially or completely blocking air as it passes from the lungs through the vocal tract. If you review the inventory of English consonants in Table 3.1 and pronounce the sounds aloud while concentrating on the place and manner of articulation, you'll perceive how the tables represent the distribution of English consonants according to their voicing, place of articulation, and manner of articulation. On the next page we describe these consonants, grouped according to their manner of articulation and characterized in terms of voicing and place of articulation. We concentrate on the consonant sounds of English and mention selected consonants in other languages.

Stops

Stops are formed when air is built up at a point in the vocal tract and released suddenly through the mouth. The principal stops of English are [p] and [b], [t] and [d], and [k] and [g]. By noticing where the air is blocked while pronouncing words with these sounds in them (see Table 3.1), you can recognize that [p] and [b] are *bilabial* stops, [t] and [d] *alveolar* stops, and [k] and [g] *velar* stops. In some positions in a word, stops in English, when released, are followed by a burst of air called aspiration, a topic to which we'll return in the next chapter. Here, let it suffice to illustrate the difference with an example: the pronunciation of [p] in *pot* has aspiration, while in *spot* it does not have aspiration. (Try it with the back of your hand in front of your mouth.)

English Stops

	PLACE OF ARTICULATION			
	Bilabial	**Alveolar**	**Velar**	**Glottal**
Voiceless	p	t	k	ʔ
Voiced	b	d	g	

In addition, many languages have a glottal stop, represented by [ʔ]. It is pronounced by using the glottis to completely but briefly block the air from passing in the throat. In English, the glottal stop occurs only as a marginal sound—for example, in American English between the two parts of the exclamation *Uh-oh!* and in Cockney English as the medial consonant of words like *butter* and *bottle*. In Hawaiian and some other languages, the glottal stop is not a marginal sound but an ordinary consonant that can distinguish between words, as in *paʔu* 'smudge' and *pau* 'finished.' (Represented by an apostrophe in its spelled name, the constructed language Na'vi, spoken in the film *Avatar*, is pronounced [nɑʔvi].)

Fricatives

Fricatives are characterized by forcing air in a continuous stream through a narrow opening at some point in the vocal tract. To pronounce the alveolar **fricatives** [s] and [z], air is forced through a narrow opening between the tip of the tongue and the alveolar ridge. In pronouncing the first sound in the words *thin*, *three*, and *theta* and the final sound in *teeth*, *bath*, and *breath*, the tongue tip is placed between the upper and lower teeth, where the airstream is most constricted and makes its articulation. Represented by [θ], the sound in these words is a voiceless interdental fricative. Its voiced counterpart is the initial sound in the words *there* and *then*, the middle consonant sound in *either*, and the final sound of *bathe* and *breathe*. In English the spelling <th> is used for two distinct sounds: [θ], as in *thin*, *ether*, and *breath*, and [ð] as in *then*, *leather*, and *smooth*.

Try It Yourself Pronounce the following words to discover other fricatives and become aware of their common properties and their different places of articulation:

fine/vine; beefish/peevish	[f]	[v]	labio-dental fricatives
thigh/thy; ether/either	[θ]	[ð]	interdental fricatives
sink/zinc; bus/buzz	[s]	[z]	alveolar fricatives
rush/rouge; fishin'/vision	[ʃ]	[ʒ]	postalveolar fricatives
here; ahoy	[h]		glottal fricative

English Fricatives

	PLACE OF ARTICULATION				
	Labio-Dental	**Interdental**	**Alveolar**	**Postalveolar**	**Glottal**
Voiceless	f	θ	s	ʃ	h
Voiced	v	ð	z	ʒ	

Some languages have other fricatives. Spanish has a voiced bilabial fricative (represented by IPA [β]), as in the <b> of *cabo* 'end.' Japanese has a voiceless bilabial fricative, represented by [ɸ] and pronounced somewhat like [f] but by bringing together both lips instead of the lower lip and the upper front teeth. The West African language Ewe has voiced [β] and voiceless [ɸ] bilabial fricatives. Spanish and many other languages have a voiceless velar fricative [x] and a voiced velar fricative [ɣ]. Pronounce [x] as if you were gently clearing your throat. The sound occurs initially in the Spanish word *joya* 'jewel' and the personal name *José* (when borrowed into English, *José* is pronounced with [h], the closest sound to [x] in English). [ɣ] is represented by <g> in Spanish *lago* 'lake.' German, Irish, and Mandarin have a voiceless palatal fricative [ç], as in the German word *Reich* 'empire.'

You may have noticed that the physical distance in the mouth between the places of articulation for the English fricatives is not as great as for the stops. The bilabial, alveolar, and velar places of articulation for stop consonants are spaced farther apart than are the labio-dental, interdental, alveolar, and postalveolar fricatives. This closer spacing of the fricatives can cause difficulty in perceiving them as distinct, especially for speakers of languages with fewer fricatives than English or with fricatives spaced at greater distance from one another. French, for example, does not have the interdental fricatives [θ] and [ð], so French speakers tend to perceive (and pronounce) English words like *thin* and *this* as though they were "sin" and "zis." One French fricative sound familiar to English speakers, even though English doesn't have it, is the voiced uvular *r*-sound (as in *Paris* or *rue* 'street'), which is represented by IPA [ʁ].

Affricates

Two consonant sounds of English are more complex than its stops and fricatives. These more complex sounds combine a stop consonant and a fricative to produce what is called an **affricate**. In the pronunciation of an affricate, air is built up by a complete closure of the oral tract at some place of articulation and then released (like a stop) and continued (like a fricative). The consonant sound at the beginning and end of *church* is a combination of the voiceless stop [t] and the voiceless fricative [ʃ] and is represented as [ʧ] or [tʃ]. The consonant sound at the beginning and end of *judge* is a combination of the voiced stop [d] and the voiced fricative [ʒ] and is represented as [ʤ]. English has only one pair of affricates, and to identify their place of articulation they are called postalveolar affricates.

Some languages have other affricates. The most common are the alveolar affricates [ʦ] and [ʣ], which occur at the beginning of the Italian words *zucchero* 'sugar' and *zona* 'zone,' respectively.

Try It Yourself Focusing on the initial consonant sounds in *choke* and *joke* and the final consonant sounds in *batch* and *badge*, pronounce these words slowly until you recognize that they begin or end with stop-fricatives (that is, affricates).

English Affricates

	PLACE OF ARTICULATION
	Postalveolar
Voiceless	ʧ
Voiced	ʤ

Obstruents

Fricatives, stops, and affricates are called **obstruents** because they share the phonetic property of constricting or obstructing the airflow through the vocal tract.

Approximants

Approximants are produced by two articulators approaching one another almost like fricatives but not coming close enough to produce friction. English has four approximants: [j], [r] (IPA [ɹ]), [l], and [w]. The sound that begins the word *you* is the palatal approximant [j], while *cute* begins with the consonant cluster [kj]. Because [r] is pronounced by channeling air through the central part of the mouth, it is called a central approximant. To pronounce [l], air is channeled on one or both sides of the tongue to make a lateral approximant. To distinguish them from the other approximants, [r] and [l] are sometimes called **liquids**. (In some Asian languages, [r] and [l] are not contrastive sounds, so native speakers of those languages may find it challenging to distinguish them in speaking or perceiving English speech. This is a matter to which we return in the next chapter.)

In pronouncing the approximant [w], the lips are rounded, as in *wild* or *wow*. In certain dialects, [h] precedes [w] in words such as *which*, *whether*, or *when*. When [w] is the second element of a consonant cluster (as in *twine* [tw] or *queen* [kw]), the initial sound (here, [t] or [k]) is rounded in anticipation of the [w], as you can appreciate by focusing on the shape of your lips while you pronounce *time* and *twine* and *keen* and *queen*.

English Approximants

	PLACE OF ARTICULATION		
	Bilabial	**Alveolar**	**Palatal**
Voiced (Central)	w	r (IPA ɹ)	j
Voiced (Lateral)		l	

Nasals

Nasal consonants are pronounced by lowering the velum, thus allowing the stream of air to pass out through the nasal cavity instead of through the oral cavity. English has three nasals: [m] as in *mad*, *drummer*, *cram*; [n] as in *new*, *sinner*, *ten*; and a third, pronounced as in the words *sing* and *singer* and symbolized by [ŋ].

English Nasals

PLACE OF ARTICULATION		
Bilabial	**Alveolar**	**Velar**
m	n	ŋ

Because of the way it is usually spelled in English words, English speakers may think of [ŋ] as a combination of [n] and [g], but it is actually a single sound. You can test this for yourself by pronouncing *finger* and *singer*. Ignoring the initial sounds [f] and [s], if your pronunciations of *finger* and *singer* differ (for some speakers of English they may not), then you have [ŋg] in *finger* and [ŋ] in *singer*. Notice that if you had [ng] (instead of [ŋg] in *finger*, you'd pronounce it like "finn-ger." Most speakers of American English make a three-way distinction among *simmer*, *sinner*, and *singer*, with middle consonants [m], [n], or [ŋ]. By noticing whether and where your tongue touches the upper part of your mouth in articulating these nasal consonants (and by comparing their place of articulation with other sounds), you can determine for yourself that [m] is a bilabial nasal, [n] an alveolar nasal, and [ŋ] a velar nasal.

If you have successfully identified the places of articulation for nasals and understood why they fit in their particular slots in the consonant table, you may

Try It Yourself While you are saying a prolonged [mmmm] or [nnnn], cut off the airstream passing through your nose by pinching it closed. You'll notice that the sound stops abruptly, thus demonstrating that air passes through the nose when you produce a nasal. Compare what happens when you cut off the air passing through your nose while saying lengthened [nnnn] and [ssss]. You'll quickly understand one way in which the nasal and oral cavities function in producing sounds. When you cut off air passing through the nose, there is little difference in the sound quality for oral consonants like [s], but the effect is plainly noticeable for nasal consonants like [n] and [m].

have noticed that English has three sets of consonants articulated in the same places and differing only in their manner of articulation: the oral stops [p] and [b] and the nasal [m] are bilabials; the oral stops [t] and [d] and the nasal [n] are articulated at the alveolar ridge and called alveolars; and [k], [g], and [ŋ] are articulated at the velum and called velars.

English, as we've seen, has three nasal consonants: [m], [n], and [ŋ]. The sound inventory of other languages may contain additional nasals. French, Spanish, and Italian speakers have a palatal nasal [ɲ]. You can recognize it in the French word *mignon* 'cute' (which English has borrowed in the phrase *filet mignon*), in the Spanish words *mañana*, *señor*, and *cañón* (the last one borrowed into English as *canyon*), and the Italian *bagno* 'bath' and *lasagna* (the latter also borrowed into English).

Clicks, Taps, Trills

Some languages of southern Africa have among their inventory of stop consonants certain **click** sounds that function as an integral part of their sound system. One example is the lateral click, made on the side of the tongue. Although the lateral click is not part of the inventory of English speech sounds, it does occur when English speakers urge a horse to move on. As a speech sound, its IPA symbol is [||]. Another click sound is made with the tip of the tongue at the teeth or the alveolar ridge and is described as a dental click [|] or (post)alveolar click [!]. (That click does not occur as an English speech sound, but you can recognize it in English writing by the reproachful noise spelled *tsk-tsk*.)

A few consonant sounds are not stops, fricatives, affricates, approximants, or nasals. The middle consonant sound in the words *butter* and *metal* is commonly pronounced in American English as an alveolar **tap**—a high-velocity short flap produced by tapping (or flapping) the tongue against the alveolar ridge. We represent this tap by [ɾ] (we'll discuss it further in the next chapter). Spanish, Italian, and Fijian have an alveolar **trill** *r*, as in Spanish *correr* 'to run.' This book and some others represent the alveolar trill by [r̃] instead of the IPA symbol [r] so as to reserve the familiar symbol [r] for representing the "r" of English.

Vowel Sounds

Vowel sounds are articulated without complete closure in the oral cavity and without sufficient narrowing to create the friction characteristic of consonants. They are produced by passing air through different shapes of the mouth, with different positions of the tongue and lips, and—unlike consonants—with the airstream relatively unobstructed by narrow passages except at the glottis. Some languages have as few as three distinct vowels; others, including English, have more than a dozen.

Vowel Height and Frontness

Vowels are characterized by the position of the tongue and the relative rounding of the lips. For that reason and partly on the basis of auditory perception, we refer to vowels as *high* or *low* and *front* or *back*. We also consider whether the lips are *rounded* (as for *boot*) or *nonrounded* (as for *bit*).

Figure 3.4 indicates the relationship of the English vowels to one another and the approximate positions of the tongue during their articulation.

At the top of the following page are English words for each of the vowel symbols shown in the figure. Note that these words are chosen on the basis of North American English and that British English pronunciations may differ for some of them. Note, too, that in some North American dialects the vowels of *bought* and *pot* are pronounced alike (see Chapter 11).

Try It Yourself You've heard people say that English has five vowels—a, e, i, o, and u. But that number better reflects written than spoken English. To determine a more accurate count, pronounce the following words aloud, noting how many distinct vowels of spoken English they represent: *peat*, *pit*, *pate*, *pet*, *pat*, *put*, *putt*, *pot*, *pout*, *poke*, *pike*, and *pool*.

Try It Yourself You can get a feel for the descriptors front and back by alternately saying *feed* and *food*—*feed* contains a front vowel, *food* a back vowel. For tongue height, alternate saying *feet* and *fat*. If you don't feel the difference between high and low vowels with this pair of sounds, look at yourself in the mirror (or at a classmate saying them); the mouth is open wider for the vowel of *fat* than for the vowel of feet. The reason? The tongue is lower for *fat*.

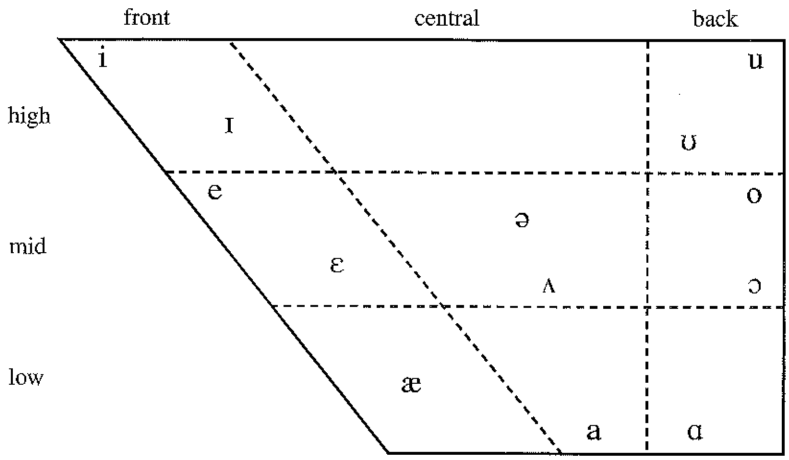


Figure 3.4 The Vowels of English

i	feet, beat			u	stoop, boot
ɪ	pit, bit			ʊ	put, foot
e	late, bait	ə	about, sofa	o	poke, boat
ɛ	pet, bet	ʌ	putt, but	ɔ	port, bought
æ	pat, bat	a	park (in Boston)	ɑ	pot, father

The symbols [ə] (called schwa) and [ʌ] (called wedge) represent similar sounds. Both occur in the word *above* [əbʌv]. We use [ə] to represent a mid-central vowel in unstressed syllables, such as the second syllable of *buses* [bʌsəz] and the second and third syllables of *capable* [kepəbəl]. We also use it before [r] in the same syllable, whether stressed as in *person* [pərsən] and *sir* [sər] or unstressed as in *pertain* [pərten] and *tender* [tɛndər]. We use [ʌ] to represent mid-central vowels in other stressed syllables, such as *suds* [sʌdz] and the first syllable of *flooded* [flʌɾəd]. (Some books use [ɚ] to represent a mid-central vowel with *r* coloring. In systems using the [ɚ] notation, *person* would be transcribed [pɚsən], *sir* [sɚ], and *pertain* [pɚten].)

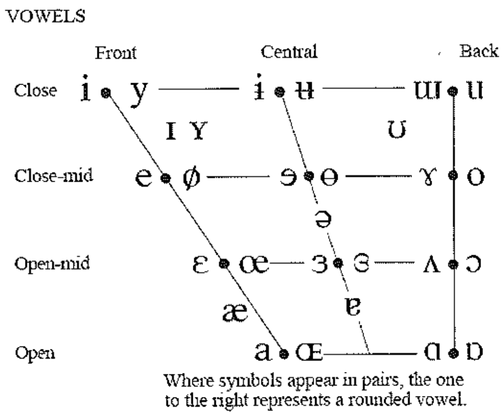


Figure 3.5 International Phonetic Association Vowel Chart

Source: IPA Chart, http://www.langsci.ucl.ac.uk/ipa/ipachart.html, available under a Creative Commons Attribution-Sharealike 3.0 Unported License. Copyright © 2005 International Phonetic Association.

Figure 3.5 shows the vowel chart of the International Phonetic Association. Most of its 28 symbols are used for the sounds of the various languages whose vowel inventories we describe later in this chapter. The IPA's use of the terms "close" and "open" parallel the use in this book of high and low, while the IPA's "close-mid" and "open-mid" parallel our use of upper mid and lower mid.

Diphthongs

English also has **diphthongs**, represented by pairs of symbols to capture the fact that a diphthong is a vowel sound for which the tongue starts in one place and glides to another. Say these slowly to get a sense of what a diphthong is: [aɪ] (as in *try* or *ride*); [aʊ] (as in *cow* or *loud*); [ɔɪ] (as in *soy* or *toyed*). You'll note a change in their quality while you pronounce them. Try again with *ply*, *plow* (or *plough*), *ploy*. (Some books transcribe these diphthongs as [ay] or [aj], [au] or [aw], and [ɔy] or [ɔj], respectively.)

Thus American English dialects have up to 13 distinctive vowel sounds (plus three diphthongs). In England and certain parts of the United States, including metropolitan New York City, 16 distinct vowels and diphthongs exist. In some North American dialects of English, speakers make no distinction between the vowels of *hawk* and *hock*, so those dialects have fewer vowel sounds.

Other Articulatory Features of Vowels

To create differences among vowels, languages can exploit other possibilities besides tongue height and frontness or backness. Vowels can have rounding, tenseness, length, nasalization, and tone.

Table 3.2 *French Vowels with Illustrative Words*

	Front Unrounded	Front Rounded	Central Unrounded	Back Rounded
Oral				
high	i	ü		u
upper mid	e	ø		o
mid			ə	
lower mid	ɛ	œ		ɔ
low			a	
Nasal				
lower mid	ɛ̃	œ̃		ɔ̃
low				ã
	i gris 'gray'	ü mûr 'ripe'	ə chemin 'path'	u fou 'crazy'
	e fermé 'shut'	ø jeûne 'fasts'	a par 'by'	o mot 'word'
	ɛ frais 'fresh'	œ jeune 'young'		ɔ fort 'strong'
	ɛ̃ brin 'sprig'	œ̃ brun 'brown'		ɔ̃ fond 'bottom'
				ã faon 'fawn'

Rounding High front vowels in English tend automatically to be unrounded (and high back vowels to be rounded) as we saw above, but some other languages have *rounded* and *unrounded* front vowels. French and German have high front and mid front rounded vowels as well as unrounded ones. French, besides having a high front unrounded [i] in words such as *dire* 'to say' and *dix* 'ten,' also has a high front rounded vowel [ü], as in *rue* 'street.' French also has a contrast between upper mid front unrounded [e] (as in *fée* 'fairy') and upper mid front rounded [ø] (*feu* 'fire') and between lower mid front unrounded [ɛ] (*serre* 'hothouse') and lower mid front rounded [œ] (*soeur* 'sister'). (See Table 3.2.) German has similar contrasts. (See Table 3.4.)

Tenseness Languages may make a distinction between *tense* and *lax* vowels. Lax vowels tend to be shorter and do not occur at the end of a stressed syllable; they also tend to be more centralized than the nearest tense vowel. The contrast between the [i] of *peat* and the [ɪ] of *pit* is in part a tense/lax contrast; likewise for the vowels in *bait/bet* and *cooed/could.* Lax vowels don't end a syllable (we have [pi] 'pea' but not [pɪ] except preceding a consonant, as in [pɪk] *pick* or [pɪt] *pit*); they are shorter than their corresponding tense vowels and more centralized in the mouth. Thus English has the lax vowels [ɪ ɛ ʊ] as in *pit, pet, put,* and they appear closer to the center on the vowel chart (see Figure 3.4), representing their greater centralization compared to their tense counterparts. The corresponding tense vowels are [i e u] as in *beat, bait, boot,* and in each case they appear more to the periphery of the vowel chart than the corresponding lax vowel. The vowels [æ ʌ] are lax and do not have corresponding tense vowels.

Table 3.3 *Spanish Vowels with Illustrative Words*

	Front Unrounded	Central Unrounded	Back Rounded
High	i		u
Mid	e		o
Low		a	
	i chiste 'joke'	a mar 'sea'	u sur 'south'
	e fe 'faith'		o boca 'mouth'

Table 3.4 *German Vowels with Illustrative Words*

	Front Unrounded	Front Rounded	Central Unrounded	Back Rounded
HIGH				
long	i:	ü:		u:
short	i	ü		u
UPPER MID				
long	e:	ø:		o:
short	e			o
MID				
short			ə	
LOWER MID				
long	ɛ:			
short		œ		
LOW				
long			a:	
short			a	
	i: bieten 'to wish'	ü: Mühle 'mill'	ə liebe 'dear'	u: Huhn 'hen'
	i bitten 'to request'	ü müssen 'must'	a: Rabe 'raven'	u Mutter 'mother'
	e: wen 'whom'	ø: ölig 'oily'	a Ratte 'rat'	o: Ofen 'oven'
	e wenn 'when'	œ Röntgen 'X-ray'		o Ochs 'ox'
	ɛ: Käse 'cheese'			

Table 3.5 *Japanese Vowels with Illustrative Words*

	Front Unrounded	Central Unrounded	Back Unrounded	Back Rounded
High	i		ɯ	
Mid	ɛ			ɔ
Low		a		
	i ima 'now'	a aki 'autumn'	ɯ buji 'safe'	ɔ yoru 'to approach'
	ɛ sensei 'teacher'			

Discrete Sounds or Overlapping Sounds

Alphabetic writing systems rely on a notion of discrete sounds, and it is useful to think of speech sounds as discrete. At the beginning of this chapter, we identified words as having a specific number of sounds: *sat* with three, *untold* with six. But in natural speech, sounds are not discrete and don't occur separately. Instead, more than touching the sounds on either side of it, each sound in a word (or an utterance) blends into them to some degree.

Imagine a computerized sound synthesizer that could create discrete sounds that seem natural when spoken in isolation—it could produce the sounds [ɪ], [n], [k], [l], [u], [d], [ə], and [d]. If it put these sounds together in the sequence [ɪnkludəd], you might expect to hear a pronunciation that resembled the word *included*. But it would not sound like *included*. As it is usually pronounced, *included* does not have the same [ɪ] sound that occurs in *sit* (for one thing, the [ɪ] of *included* is nasalized). Further, the [n] of *included* is often pronounced more like the [ŋ] of *sing* as the vocal tract positions itself in anticipation of the velar [k] sound that follows. So if natural-sounding words are the goal, their production cannot rely on a sequence of discrete individual sounds.

Even a computerized sequence of [ɪ̃], [ŋ], [k], [l], [u], [d], [ə], and [d] would not sound natural. For one thing, the first and second [d] sounds of *included* are different from one another (and neither of them is just like the [d] sound of *dig*). Further, the individual sounds would have to run into one another as in natural speech. They couldn't be separated as in print or a phonetic transcription or beads on a string. In other words, without some general principles of pronunciation, a computer could not simply combine the sounds represented in spelling or even a phonetic transcription and produce synthesized speech that sounded like natural speech. Still, we know from everyday encounters on the telephone when checking a bank account balance or making a directory inquiry to find someone's telephone number or establishing an airplane reservation that computerized speech often sounds surprisingly natural. How does that happen? We return to that question in the next chapter.

Length Vowels can be of varying lengths. Long vowels are held longer when pronounced than short vowels. In phonetic transcriptions a special colon or doubled vowel symbol may be used to indicate a long vowel. (In dictionaries and some writing systems, a macron (¯) above the vowel symbol may indicate length.) German is a language where each vowel has a contrasting *long* and *short* version. Thus, in addition to the short vowels [i] and [ü], as in *bitten* 'to request' and *müssen* 'must,' German has words with high front long vowels: unrounded [i:] in *bieten* 'to wish' and rounded [ü:] in *Mühle* 'mill.' These examples illustrate how languages can multiply vowel differences by exploiting length. To sense differences in the duration of vowels, pronounce the English words *bit, beat,* and *bead.* You should be able to hear that the vowel of *bead* is held longer than the vowel of *beat,* and that both are held longer than the vowel of *bit.*

Nasalization All vowel types can be *nasalized* by pronouncing the vowel while passing air through the nose and through the mouth. The IPA marks nasal vowels by placing a tilde [˜] above the vowel symbol. French has several nasal vowels paralleling its oral vowels:

lin [lɛ̃] 'flax'	*lait* [lɛ] 'milk'
ment [mɑ̃] '(he) is lying'	*ma* [mɑ] 'my' (feminine)
honte [ɔ̃t] 'shame'	*hotte* [ɔt] 'hutch'

Irish, Hindi, and the Native American languages Delaware, Navajo, and Seneca also have nasal vowels.

Tone In many languages of Asia, Africa, and North America, a vowel can be pronounced on several pitches and will be perceived by native speakers of those languages as different sounds. Typically, a vowel pronounced on a low pitch contrasts with the same vowel pronounced on a higher pitch. In Hausa, spoken in West Africa, the word for 'bamboo' is *górà*, with high tone (′) on the first syllable and low tone (`) on the second syllable. If the tones are reversed as in the word *gòrá*, the meaning is 'large gourd.' Some languages have more complex **tone** systems. Mandarin Chinese has a high level tone (symbolized with ¯); a rising tone (′); a falling-rising tone (ˇ), in which the pitch begins to fall and then rises sharply; and a falling tone (`), in which the pitch falls sharply. There is a four-way contrast in tone among the following vowels, which makes them independent and different words.

ī (high level)	'one'
í (rising)	'proper'
ǐ (falling-rising)	'already'
ì (falling)	'thought'

While tone is a linguistic notion, pitch refers to a physical sound and, because pitch ranges differ between individuals (for example, men have characteristically lower pitch ranges than women) a hearer will perceive the relative differences in a single speaker's pitch range as the same pattern of linguistic tone as the relative differences in another speaker's pitch range. It is the levels of pitch relative to one another within a speaker's range that will be interpreted by a hearer as signaling various linguistic tones. To put it another way, the physical characteristics of a sound wave created by speaking may differ from person to person but will be perceived by fellow speakers as signaling the same linguistic content.

Figure 3.6 is a graphic representation of pitch contrasts for the four Mandarin words listed above as they were pronounced by an adult male native speaker of Mandarin. The left-most line—for high level tone ī—is relatively high and flat compared to the others. The second pitch contour to the right—for rising tone í—starts relatively low and rises. The third contour—for falling-rising ǐ—initially falls somewhat and then rises dramatically. The right-most pitch contour—for falling ì—begins high and falls steeply. Because the values for these different tone pitches are not important in absolute terms, we haven't indicated the pitch frequencies in the figure. You needn't be interested in the physics of these particular tones for our purposes here, but just in case you

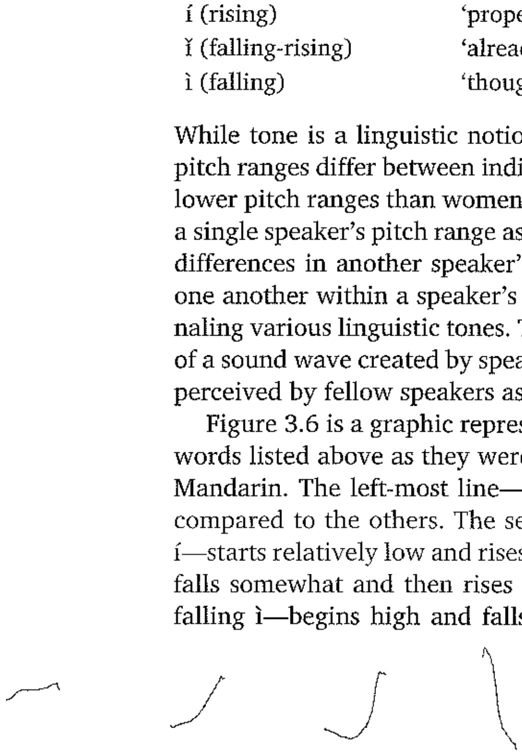


Figure 3.6 Four Tones in Mandarin Chinese

wish to know, the fourth contour falls from about 220 Hz to 60 Hz (Hz stands for Hertz and is the measure of a sound wave's frequency in cycles per second).

A given **diacritic**, or accent mark, may represent different tones in different languages. Thus, ´ represents a high tone in Hausa but a rising tone in Mandarin, while ` represents a low tone in Hausa but a falling tone in Mandarin.

Try It Yourself Figure 3.6 was made using the Praat software program (freely available for downloading at <www.praat.org>). Using Praat, you will be able to record and examine a spectrogram of these four tones with the help of a Mandarin-speaking classmate. The patterns will look like those shown in the figure.

Thai has five tones; the standard dialect of Vietnamese has six tones; and the Guangzhou (Canton) dialect of Chinese has nine tones. Tone, as you can see, is a widespread and diverse phenomenon.

Tables 3.2 through 3.5 are vowel charts illustrating the sound patterns of French, Spanish, German, and Japanese.

Table 3.6 summarizes all the consonant sounds introduced in this chapter, while Figure 3.5 summarizes all the vowel sounds.

Table 3.6 *Consonants Discussed in Chapter 3*

	PLACE OF ARTICULATION								
Manner of Articulation and Voicing	**Bilabial**	**Labioden-tal**	**Inter-dental**	**Alveolar**	**Postal-veolar**	**Palatal**	**Velar**	**Uvular**	**Glottal**
STOPS									
voiceless	p			t			k		ʔ
voiced	b			d			g		
NASALS									
	m			n		ɲ	ŋ		
FRICATIVES									
voiceless	ɸ	f	θ	s	ʃ	ç	x		h
voiced	β	v	ð	z	ʒ		ɣ	ʁ	
AFFRICATES									
voiceless				ts	tʃ				
voiced				ʣ	ʤ				
APPROXIMANTS									
voiced central	w			r (IPA ɹ)		j			
voiced lateral				l					
OTHERS									
voiced trill				r̃ (IPA r)					
voiced tap				ɾ					

Computers and Phonetics

Who Speaks for Watson? Who Listens for Watson? In early 2011, IBM's computer named Watson competed against two human champions on the popular television game show *Jeopardy!* To maximize their winnings, as you may know, contestants must speedily and correctly offer a question whose answer is the prompt just presented by the host. Among the prompts challenging these contestants was this one: "While Maltese borrows many words from Italian, it developed from a dialect of this Semitic language." "What is Arabic?" Watson immediately replied. To the prompt "Aeolic, spoken in ancient times, was a dialect of this," Watson also replied correctly: "What is Ancient Greek?" To the prompt "Vedic, dating back at least 4,000 years, is the earliest dialect of this classical language of India," Watson knew the correct question: "What is Sanskrit?" Racking up impressive wins over the course of three days, Watson outperformed both of the human champions and spoke his replies with a high degree of naturalness, though not enough to hide the fact that his speech was synthesized. The computer gave voice to its answers using a sophisticated text-to-speech synthesizer that had been improved greatly by IBM in the year leading up to the much anticipated contest between human and machine.

Greatly impressed with Watson's performance, most television viewers and some journalists likely believed that Watson was responding to host Alex Trebek's spoken prompts. Watson could speak and was widely thought to recognize and understand speech as well. In fact, Watson could *not* recognize speech! Instead, when a written question was revealed to contestants and read aloud by Alex Trebek, Watson received an instantaneous ASCII-file formatted version of the written prompt. A written version of the prompt is visible to contestants while the host reads it aloud, and attempted answers are not possible until the prompt has been fully voiced.

Producing synthesized speech from written text (speech synthesis) and producing written text from spoken language (speech recognition) are complex enterprises, and great progress has been made in both directions. Software reflecting impressive speech synthesis and speech recognition capabilities is increasingly available in the marketplace. Apple's Siri received a good deal of attention when it was released, and Apple boasted in its advertising that "Siri not only understands what you say, it's smart enough to know what you mean." Competitors offer systems of their own. But don't be fooled into thinking that speech recognition and speech synthesis are straightforward or fully developed. The technology and language analysis that lie behind these software systems are enormously complex and have taken more than a century to develop. ■

Summary

- Sounds must be distinguished from letters and other visual representations of language.
- Phonetic alphabets represent sounds in a way that is consistent and comparable across different languages; each sound is assigned a distinct representation, independent of the customary system used to write a particular language.
- With few exceptions, this chapter uses the International Phonetic Alphabet (IPA) to represent sounds.
- All languages contain consonants and vowels.
- Consonants are produced by obstructing the flow of air as it passes from the lungs through the vocal tract and out through the mouth or nose.
- For fricative consonants, air forced through a narrow opening forms a continuous noise, as in the initial and final sounds of *says* [sɛz] and *fish* [fɪʃ].

- For stop consonants, the air passage is completely blocked and then released, as in the initial and final sounds of the words *tap* and *cat*.
- Affricates are produced by combining a stop and a fricative, as in the initial sound of *cheat* and the final sound of *peach* and the initial and final sounds of *judge*.
- As a group, fricatives, stops, and affricates are called obstruents.
- An approximant is produced when one articulator approaches another but the vocal tract is not sufficiently narrowed to create the audible friction of a consonant, as in the initial sounds of *west* [wɛst], *yes* [jɛs], *rest* [rɛst], *lest* [lɛst].
- "Liquid" is a cover term for [r] and [l] sounds.
- Consonant sounds can be described as a combination of articulatory features: voicing, place of articulation, and manner of articulation. For example, [t] is a voiceless alveolar stop and [v] a voiced labio-dental fricative.
- Vowels are produced by positioning the tongue and mouth to form differently shaped chambers and passages.
- The airstream for oral vowels passes through the mouth; for nasal vowels, the airstream passes through the nose or the nose and mouth.
- Vowels are described by relative height and relative frontness in the mouth. For example, [æ] is a low front vowel and [u] a high back vowel.
- Secondary features of vowel production—such as length, nasality, tenseness, or rounding—are sometimes specified, as in "long vowel" or "nasal vowel."
- In many languages, vowels can be pronounced on different pitches, or tones.
- Languages differ from one another in the number of speech sounds in their inventories.
- Although linguists find it useful to conceptualize the sounds of speech as separate and discrete from one another, the sounds of natural speech are connected and overlapping.

What Do You Think? REVISITED

1. *How many vowels?* An easy way to figure out which vowels exist in English is to take a simple word frame like *b_t* and see how many different vowels you can set inside to produce a different word: *bit, beet, bet, bait, bat, but, boot, boat, bought, bot, bite, bout*. One vowel that doesn't occur in that frame but occurs in the *p_t* frame is *put*. That's already 13 vowels, far more than the five suggested by those in the alphabet.
2. *Sounds and letters*. Ron rejects <ph> for *physics* because an <f> spelling is available. But he can't object to <th> for the initial sound in *then* and final sound in *smooth*; <sh> for the initial sound in *ship* and final sound in *wish*; <ch> as the initial sound of *chair*, the final sound of *which*, and the initial and final sounds of *church*.
3. *Put and putt*. English words that are pronounced and spelled differently but whose spelling difference doesn't correspond to the pronunciation difference include *satin/Satan*; *bit/bite*; *lit/light*; *woman/women*.
4. *Organs of the vocal tract*. Bio major Bill is right. The organs we use to produce human speech evolved to serve more basic functions, including eating and breathing. Speech has likely evolved to take advantage of the organs available for those basic functions.

Exercises

Practice Exercise

The words in each minimal pair below differ from one another in only a single consonant, and within each pair the different consonant sounds differ in only one or two properties (voicing, manner of articulation, or place of articulation). For each pair, give the IPA symbols for the contrasting consonant sounds and then identify the properties that differentiate the consonants.

Examples: fat/vat: f/v (voicing); vat/that: v/ð (place of articulation); wren/red: n/d (manner of articulation)

sin/sing	either/ether
pit/bit	arrive/arise
pit/pick	thief/fief
dig/gig	chief/sheaf
ten/den	rung/young
Sam/sham	rung/rum
shirk/jerk	climb/crime
many/penny	farm/charm

Based on English

3-1. Referring to Figure 3.5 and Table 3.6 or the inside back cover, give a phonetic description of the sounds listed below. For consonants, include voicing and place and manner of articulation. For vowels, include height, a frontness/backness dimension, and (where needed) a tense/lax distinction. *Examples*: [s]—voiceless alveolar fricative; [i]—high (or close) front tense vowel

Consonants: [z] [t] [b] [n] [ŋ] [r] [j] [ʃ] [θ] [ð]
Vowels: [ɛ] [æ] [ɔ] [ɪ] [ʊ] [o] [ə] [ɑ] [e] [aɪ]

3-2. A pair of words that have the same sounds in the same order, except that a single sound differs, is called a minimal pair, as in these examples: *pit* [pɪt] / *bit* [bɪt]; *bell* [bɛl] / *bill* [bɪl]; and *either* [iðər] / *ether* [iθər].

a. For each of the following pairs of English consonants, provide minimal pairs that illustrate their occurrence in initial, medial, and final position. (Examples are given for the first pair.)

		Initial	Medial	Final
[s]	[z]	sue/zoo	buses/buzzes	peace/peas
[k]	[b]	________	________	________
[t]	[b]	________	________	________
[s]	[t]	________	________	________
[r]	[l]	________	________	________
[m]	[n]	________	________	________

b. For each vowel pair below, cite a minimal pair of words illustrating the contrast.

Example: [u] [æ] *boot/bat.*

[i] [ɪ]; [ɔɪ] [aɪ]; [u] [ʊ]; [æ] [e]

3-3. Write out in ordinary spelling the words represented by the following transcriptions. *Examples*: [pɛn] *pen*; [smok] *smoke*; [bənænə] *banana*

[ton]	[plæɾər]	[læŋgwədʒ]	[træpt]	[spaʊts]
[tʃip]	[prɛpəreʃən]	[θwɔrt]	[ðiz]	[ðɪs]
[dʒeɾəd]	[ʌðərwaɪz]	[lʌvd]	[plɛʒər]	[kwɪkli]
[ɑnər]	[feɾəlɪstək]	[mənɑgəməs]	[frənɛɾək]	[ɛntərpraɪzɪŋ]

3-4. Below are phonetic transcriptions of the names of popular movies. Write the names using ordinary English spellings. *Example*: for [bʌɾi] you would write Buddy; for [ðə dɑrk naɪt raɪzəz], The Dark Knight Rises.

[ɑrgo]
[ævətɑr]
[nɔrθ kʌntri]
[dʒeŋgo əntʃend]
[mɑrvəlz ði əvɛndʒərz]
[praɪd ən prɛdʒədəs]
[ðə hʌŋgər gemz]
[gʊd naɪt ən gʊd lʌk]
[ælvən ən ðə tʃɪpmʌŋks tʃɪprɛkt]
[ʃərlɑk homz ə gem əv ʃædoz]
[slʌmdɔg mɪljənɛr]
[ði əmezɪŋ spaɪɾər mæn]
[hæri pɑɾər ən ðə hæf blʌd prɪns]
[tʃɑrli ən ðə tʃɑklət fæktri / tʃɔklət]
[ðə kjuriəs kes əv bɛndʒəmən bʌʔən]
[ðə krɑnəkəlz əv nɑrniə ðə laɪən ðə wɪtʃ ən ðə wɑrdrob]

3-5. For each word below a broad phonetic transcription is provided, each transcription containing at least one clear error. Identify the mistakes and correct them.

spitting	[spɪtiŋ]
trees	[tris]
Spain	[spaɪn]
scientific	[saɪəntifək]
cutting	[kʊɾɪŋg]
sketchy	[sketʃi]
women	[wɪmɛn]
psychological	[psaɪkəlɑdʒəkəl]
cuddle	[kudəl]
television	[teləvɪʒən]

3-6. The transcription below represents one person's reading of a passage about the actor Will Smith (adapted from *Newsweek*, July 7, 1997). The transcription does not represent secondary features such as vowel length or consonant aspiration, and capitalization and punctuation are not represented. Write out the passage using ordinary English spellings, as indicated in the first few lines and the last line.

wɪl smɪθ hæz ə dɑrk ferəl flɔ	Will Smith has a dark, fatal flaw.
ɪts ən əbsɛʃən əv sɔrts	It's an obsession of sorts,
ðə kaɪnd əv θɪŋ ðæt kən draɪv lʌvd	the kind of thing that can drive loved
wʊnz krezi	ones crazy.
ən maɪt ivən ɪf əlaʊd tə rʌn əmʌk	
direl ən dəbɪlətet ən ʌðərwaɪz prɑməsɪŋ kərir	
hi hets bæd græmər	
prənʌnsieʃən ɛrərz mɪsteks əv ɛni lɪŋgwɪstək sɔrt	
ðe mek hɪm nʌts	
hɪz gərlfrɛnd ði æktrəs ʤedə pɪŋkət noz ət	
əkeʒənəli ɪn ðɛr ʤɛntləst most kærɪŋ we	
ðe traɪ tə kɔʃən hɪm əbaʊt ðə sɪriəsnəs əv hɪz əflɪkʃən	
sɪrɪŋ daʊn ovər brɛkfəst wʊn mɔrnɪŋ	
ɪn ðɛr spænɪʃ staɪəl vɪlə aʊtsaɪd ɛle	
pɪŋkət kæsts ə tɛnərəv glæns ɪn hɪz dərɛkʃən	
wɑt wər jə tɛlɪŋ mi ði ʌðər de ʃi sɛz	
ðæt pipəl se ðə wərd ɔfən laɪk ɔf fən wɛn ɪts rili prənaʊnst ɔf tən	
smɪθ lʊkɪŋ spɔrri ən prɑpər in ə waɪt rælf lɔrɛn polo ʃərt	
waɪt swɛtpænts ən naɪki ɛr ʌp tɛmpoz	
sɛts daʊn ə plærər əv bənænə pænkeks wɪθ ə dɪsəpruvɪŋ θʌd	
no no hi sɛz	
ðə raɪt we ɪz ɔfən	
pipəl hu prənaʊns ðə ti ɑr traɪɪŋ tə saʊn səfɪstəkerəd	
bət ðe ʤʌst saʊn rɔŋ	
pɪŋkət gɪgəlz ðɛn əfɛks ə supərmæn ton əv vɔɪs	
ɪts ə naʊn ɪts ə vərb	
no ɪts kæptən kərɛkʃən	No, it's Captain Correction.

3-7. The following transcription represents one person's reading of a passage about love potions (adapted from *The Encyclopedia of Things That Never Were*, p. 159). Write out the passage using ordinary English spellings.

æz ðə nem ɪndəkets	As the name indicates,
ðiz poʃənz ɑr kɑmpaʊndəd spəsɪfəkli	these potions are compounded specifically
tu ətrækt ə sʌbʤɛkt	to attract a subject
hu ɪz rilʌktənt tə sərɛndər	who is reluctant…
tə wʊnz kɑrnəl dəzaɪərz	
ðə poʃən me bi hæd ærə praɪs frəm ɛni ælkəmɪst	

ɔr ʌðər pərsən skɪld ɪn ðə prɛpəreʃən əv maɪn tʃendʒɪŋ kɑmpaʊnz
wɪtʃəz wɪzərdz ən sɔrsərərz
hu ɑr dʒɛnrəli nɑt ɪntrəstəd ɪn lʌv
ɑr sʌmtaɪmz ənwɪlɪŋ tə mænjəfæktʃər ðə poʃənz
ðə pərtʃəsərz onli prɑbləm me bi ðæt əv pərswerɪŋ
ði ɑbdʒɛkt əv hɪz ɔr hər dəzaɪər
tə swɑlo ɛni əv ðə poʃən
ə risənt rɛsəpi fɔr ə lʌv poʃən ɪŋklurəd
dʒɪndʒər sɪnəmən draɪd ən graʊnd grep sidz
ɔɪstər ɛlk æntlər ən tel her frəm ə mel ænəməl
ænd ɛni surəbəl ɑbdʒɛkt frəm ðə pərsən
sʌtʃ æz hɪz ɔr hər nel klɪpɪŋz

3-8. Transcribe each of the following words as you say them in casual speech. (Don't be misled by the spelling; it could be helpful to have someone else pronounce them for you.) *Examples*: *bed* [bɛd]; *rancid* [rænsəd]; *shnook* [ʃnʊk]

changes	mostly	very	friend	teacher
semantics	system	ready	more	musician
crackers	peanuts	palm	music	photographer
pneumonia	attitude	psalm	fuel	photograph

3-9. Examine the following list of consonants as they are represented in four popular dictionaries (the first three are American; COD is British), and compare the dictionary symbols with the IPA symbols. The abbreviation MWCD stands for *Merriam-Webster's Collegiate Dictionary*, eleventh edition; WNWCD for *Webster's New World College Dictionary*, fourth edition; AHD for *The American Heritage Dictionary of the English Language*, fifth edition; COD for *The Concise Oxford Dictionary of Current English*, tenth edition.

IPA Symbol	MWCD	WNWCD	AHD	COD
p	p	p	p	p
k	k	k	k	k
θ	th	th	th	θ
ð	t͟h	*th*	*th*	ð
s	s	s	s	s
ʃ	sh	sh	sh	ʃ
ʒ	zh	zh	zh	ʒ
tʃ	ch	ch	ch	tʃ
dʒ	j	j	j	dʒ
ŋ	ŋ	ŋ	ng	ŋ
h	h	h	h	h
j	y	y	y	j

Some symbols used by dictionaries are the same as the IPA symbols, but not all. North American dictionaries tend to prefer their own symbols, while the

British COD leans strongly toward the IPA. Choose three sounds for which at least one dictionary uses a different symbol from the IPA symbol, and discuss why it might have been chosen.

3-10. Examine the following list of vowels as they are represented in three dictionaries; compare the dictionary symbols with the IPA symbols. (See Exercise 3-9 for identification of the dictionaries.)

IPA Symbol	Words	MWCD	WNWCD	AHD	COD
i	peat, feet	ē	ē	ē	iː
ɪ	pit, bit	i	i	ĭ	ɪ
ɛ	pet, bet	e	e	ĕ	ɛ
e	wait, late	ā	ā	ā	eɪ
æ	pat, bat	a	a	ă	a
ə	sod<u>a</u>, it<u>e</u>m	ə	ə	ə	ə
ʌ	putt, love	ə	u	ŭ	ʌ
u	pool, boot	ü	o͞o	o͞o	uː
ʊ	push, put	u̇	oo	o͝o	ʊ
o	boat, sold	ō	o	ō	əʊ
ɔ	port, or	ȯ	ô	ô	ɔː
ɑ	pot, bottle	ä	ä	ŏ	ɒ
aʊ	cow, pout	au̇	ou	ou	aʊ
aɪ	buy, tight	ī	ī	ī	ʌɪ
ɔɪ	boy, toil	ȯi	oi	oi	ɔɪ

In contrast to their practice with consonants, desk dictionaries differ noticeably from one another and from the IPA in representing vowel sounds. Cite three instances of a difference between a dictionary's symbol and that of the IPA, and discuss the advantages and disadvantages of the dictionary's representation as compared to the IPA's.

3-11. George Bernard Shaw's tongue-in-cheek claim that English spelling is so chaotic that *ghoti* could be pronounced [fɪʃ] "fish" has been called misleading. That judgment is based on observations like these: <gh> can occur word initially in only a few words (for example, *ghost* and *ghastly*), and then it is always pronounced [g]; only following a vowel in the same syllable (as in *cough* and *tough*) can <gh> be pronounced as [f]; thus, *ghoti* could not be pronounced with an initial [f]. What other generalizations about the English spelling patterns of <gh>, <o>, and <ti> can be used to argue that Shaw's claim is exaggerated?

Especially for Educators and Future Teachers

3-12. Your ESL class complains that English spelling is chaotic. Reading would be easier, they say, if spelling reflected pronunciation. As examples, they claim that *electricity* should be spelled <elektrisity> or <alektrisatee> and *electrical* <elektrikal> or <alektrakal>; likewise, they say, *cats* should be spelled <kats> and *dogs* <dogz>. In what sense could your students' claim be right? On the other hand, what arguments could you offer in support of the view that reading is easier with little or no variation in the spelling of the ELECTRIC morpheme and the 'PLURAL' morpheme even when the pronunciation differs? In other words, what good arguments can you offer for keeping traditional spellings in such cases?

3-13. As a follow-up to the discussion about consistent spelling of the same morpheme in different words even when the pronunciation of that morpheme differs in the words, you realize that native speakers from different regions have different vowel pronunciations. Some have the same pronunciation for *talk* and *tock* (and *walk* and *wok*), while others do not pronounce these pairs alike. What would your student spelling reformers propose to accommodate these pronunciation differences across different groups of speakers?

3-14. Your ESL class notices that you pronounce words like *later*, *fatter*, and *metal* as though they were spelled with <d> instead of <t>—you pronounce them as in *lady*, *ladder*, and *medal*. They ask why you don't pronounce them with the [t] sound of the spelling. What's your explanation?

Other Resources

Internet

- **LISU Website: http://www.CengageBrain.com** For users of this textbook. Provides updated Internet links as well as supplemental material for students and instructors. Here you will find interactive learning tools.
- **International Phonetic Association: http://www.langsci.ucl.ac.uk/ipa/** Here you'll find the most up-to-date version of the IPA, including vowels, consonants, diacritics, suprasegmentals, tones, and word accents. You'll also find links to sites where you can download IPA fonts for your word processing programs, as well as information about recordings of the sounds of the IPA.
- **UCLA Phonetics Lab Data: http://www.phonetics.ucla.edu/course/chapter1/chapter1.html** Here you'll be able to hear all the speech sounds represented in the IPA charts.
- **The Sounds of the IPA: http://www.phon.ucl.ac.uk/home/wells/cassette.htm** A cassette and CD of the sounds of the International Phonetic Alphabet are available. For ordering information, use the link at the IPA home page or go directly to this website.
- **Signal Analysis and Interpretation Laboratory: http://sail.usc.edu/span/video.php** At this site you'll find fascinating MRI video clips of people pronouncing many sounds of American English and Indian English. Images of the tongue and its movement, as well as other features of the vocal tract pronouncing various sounds.

Suggestions for Further Reading

- **Michael Ashby & John Maidment. 2005.** ***Introducing Phonetic Science*** (Cambridge, UK: Cambridge University Press). A basic treatment with separate chapters on voice, place of articulation, manner of articulation, suprasegmentals (like tone), and speaker and hearer; also a good introductory treatment of instrumental phonetics.
- **Dani Byrd & Toben H. Mintz. 2010.** ***Discovering Speech, Words, and Mind*** (Chichester: Wiley-Blackwell). An accessible, richly illustrated, and accessible book exploring the speech production and perception and the relationship of speech to language and mind.
- **David Crystal. 2008.** ***A Dictionary of Linguistics and Phonetics,* 6th ed.** (Oxford: Wiley-Blackwell). A helpful source of information about the meanings of terms.
- **Peter B. Denes & Elliot N. Pinson. 1993.** ***The Speech Chain: The Physics and Biology of Spoken Language,* 2nd ed.** (New York: Freeman). An accessible account of the physics and biology of spoken language; includes chapters on acoustic phonetics, digital processing of speech sounds, speech synthesis, and automatic speech recognition.
- **Peter Ladefoged. 2012.** ***Vowels and Consonants: An Introduction to the Sounds of Language,* 3rd ed., revised by Sandra Ferrari Disner** (Malden, MA: Wiley-Blackwell). Offers clear explanations of the articulatory characteristics of vowels and consonants, with accessible chapters on acoustic aspects of speech and on talking and listening computers. Highly recommended.
- **Peter Ladefoged & Keith Johnson. 2010.** ***A Course in Phonetics,* 6th ed.** (Boston: Wadsworth). An excellent introduction to the production mechanisms of speech and the variety of sounds in languages.
- **Peter Ladefoged & Ian Maddieson. 1996.** ***The Sounds of the World's Languages*** (Malden, MA: Blackwell). An advanced treatment of the articulatory and acoustic phonetics of the various sounds in the languages of the world.
- **Ian R. A. MacKay. 1991.** ***Phonetics: The Science of Speech Production,* 2nd ed.** (Boston: Allyn and Bacon). A complete elementary treatment of all aspects of phonetics; accessible with excellent illustrations.
- **Ian Maddieson. 1984.** ***Patterns of Sound*** (Cambridge, UK: Cambridge University Press). An inventory of the sounds in a representative sample of the world's languages; the inventories vary from a low of 11 to a high of 141 sounds.
- **Geoffrey K. Pullum & William A. Ladusaw. 1996.** ***Phonetic Symbol Guide,* 2nd ed.** (Chicago: University of Chicago Press). Discusses the various symbols used in the International Phonetic Alphabet (IPA) and by other writers in their treatments of phonetics and phonology; arranged like a dictionary, with each symbol clearly illustrated.
- **Michael Stubbs. 1980.** ***Language and Literacy: The Sociolinguistics of Reading and Writing*** (London: Routledge). Contains an excellent discussion of the relationship between sounds and spelling in English and other languages; offers insights into the problems facing spelling reform.

4

Sound Systems of Language: Phonology

What Do You Think?

- On a visit to Paris with your cousin Kristen, you witness her carrying on a lively conversation in French with a taxi driver taking the two of you to your hotel—and you're impressed with her. Then, as you arrive at the hotel, the driver asks Kristen whether she's Canadian or American. You can see she's crestfallen, and because she knows you've taken a linguistics course, she later asks you which features of her pronunciation betrayed her identity as an English speaker and why she hasn't been able to eliminate them. What do you say?
- The office receptionist at work hands you a written message that reads "Call Jules Biker." You're puzzled because you don't think you know anyone by that name. Aiming to jar your memory, you say "Jules Biker" aloud several times and the penny falls: it refers to your friend Jewel Spiker. When you tell Jewel about the misspelling, she can't imagine what accounts for the receptionist's perceiving *p* as *b*. Can you explain it?
- Techie friend Tammy claims machines nowadays can synthesize speech so well that it's impossible to tell the difference between a machine and a real human being. Impressed with voiced answers to questions your friends have asked their smartphones, you're less skeptical about Tammy's claim than you would have been a few years before, so you resolve to explore the subject. After 20 minutes on the Internet, what do you discover about the quality of today's speech synthesis?

- Classmates Simona from Seville and Pablo from Puebla speak Spanish natively and their English grammar and pronunciation are generally excellent. So you are puzzled by their pronunciation of words like *speak* and *study* as though they were spelled "espeak" and "estudy": they say [ɛspik] and [ɛstʌdi]. You are curious about the Spanish sound system and what prompts those pronunciations of English words. You investigate and find what?

Introduction: Sounds in the Mind

This chapter focuses on the systematic nature of speech sounds in languages. It examines which phonetic distinctions are significant enough to signal differences in meaning in a language, the relationship between how sounds are stored in the brain and how they're pronounced in a language, the ways in which sounds are organized within words in a language, and the influence neighboring sounds have on one another in pronouncing words in a language.

To appreciate the aspects of sound systems analyzed in this chapter, you'll find it useful to bear two important points in mind as you read. The first is that words need to be readily pronounceable. Although that point may seem obvious (after all, pronunciation can't exceed the limits imposed by the human vocal apparatus), what is less obvious and sometimes a challenge to realize when you think about the sounds in a word is that the vocal apparatus must move continuously from one sound to the next as a speaker talks. The second point, just as obvious but easy to overlook in language, is that listeners need to perceive different words as different. To put it another way, listeners need to recognize contrasts between one sound and another. If different sounds aren't easily recognized as different, hearers won't be able to tell readily enough what a speaker is attempting to say. To take a simple example, if you say, "I want new shoes" and your addressee hears, "I want you to choose," confusion will arise. Likewise for "Don't reply to him" and "Don't rely on him," "Don't strategize" and "Don't stretch your eyes."

Try It Yourself Think of a statement made to you or a question asked of you that you misunderstood to mean something altogether different. If you can't remember one, try to imagine such a pair.

Unless words and utterances are heard as distinct from one another, understanding another speaker would require mental telepathy. If very young children could be understood as meaning the many different things they intend when they say only "bah," what motivation would they have to learn their language—in other words, to differentiate among words? As we'll see, much of what happens in the sound patterns of every language seems to be influenced by those two simple facts: words must be easy to pronounce and words spoken must be heard as distinct.

It's useful to approach sound systems from the point of view of children acquiring a native language. Imagine the task of an infant listening to utterances made by parents, siblings, and others. From the barrage of utterances encountered in early life, a child must decipher a code and learn to speak his or her mother tongue. In addressing children, caregivers in some cultures use slow and careful speech, sometimes called "baby talk," but they don't do it consistently, and not all cultures follow that practice anyway. To make matters more difficult, the utterances that children do hear are often incomplete, interrupted, or flawed in other ways. Still, children learn their mother tongue and they learn it quickly and efficiently. (And bear in mind that children—like the rest of us—must learn that words produced by one person, though very different physically from words produced by another person—are nevertheless the same word. In addition, what language a child hears is often heard in the midst of other people talking to one another but not to the child or with a television or radio in the background, with still other people speaking. So focusing on a particular stream of speech must be achieved, too.)

In Chapter 3 we distinguished between the number of letters in a written word and the number of sound segments in its pronunciation. We took it for granted that words contain a specific number of sound segments—three in *pot*, four in *plot*, and five in *plots*. But in their early months children would seem to have no ready access to that information. Attempting to separate words from one another, never mind sounds, in even a few seconds of a conversation or radio broadcast in an unfamiliar language demonstrates how difficult a challenge it is because words run together in utterances in every language.

Wordsprintedwithoutspacesbetweenthemareprettytoughtoread, aren't they?

As you recognize, sorting out the individual words from one another would not be easy. Actually, the task is even more difficult than the run-together words in the printed sentence might suggest because the letters in the sentence above are discrete and separated from one another (*w*, then *o*, then *r*), but in speech the individual sound segments blend together into a continuous stream. To take the analogy one step further, imagine attempting to spot the beginning and end points of each letter in a handwritten sample: this imaginary challenge more closely reflects the one infants face in deciphering the code of distinctive sounds in their language. Consider the following:

In cursive writing the letters of each word are joined together.

Although anyone who knows English and the alphabet and can decipher this handwriting could count the letters in each word, there isn't any clear separation between letters in most of those words. Within a word, letters blend into one another, and in speech, sounds blend into one another: there is no separation between the individual sounds of a spoken word, and typically no separation between one word and the one following it—no beginning and no end for most of the individual sounds in the speech stream. Children nevertheless learn the words and sounds of their language with astonishing ease.

Physical Sounds and Sound Spectrograms

If you examine a physical "picture" of a word as made by a sound spectrogram, as in Figure 4.1, you can see that there is no separation between the sounds. One reason for the continuity between sound segments is that a segment's phonetic features—for example, voicing and nasalization—do not all begin or end simultaneously.

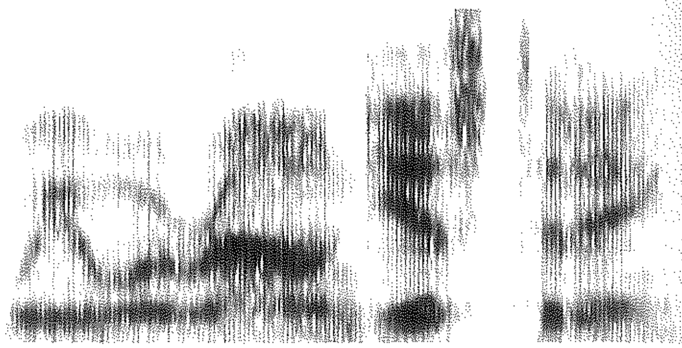


Figure 4.1 Sound Spectrogram of Utterance: "Weren't you here yesterday?"

Try It Yourself Say the words *tin* and *twin* aloud, and you'll notice that your lips are shaped very differently for the /t/ sound that begins the words. To pronounce *twin* the lips are rounded, but not when you say *tin*. Practice saying those two words until the reason for the difference becomes clear. What is it? Try also *tin* and *tune* and *kit* and *quit*.

To say the word *twin*, you don't say [t]; then [w]; then [ɪ]; then [n]. Instead, individual phonetic features of an upcoming segment are anticipated in pronouncing a preceding segment, and features of a segment may be continued into the next segment. For example, the rounding of the lips that characterizes the pronunciation of [w] affects the character of the preceding [t] in *twin*. And the vowel [ɪ], in anticipation of the nasal [n] ahead, is itself nasalized.

Similarly, the voicing feature of a voiced sound may be discontinued in anticipation of a following voiceless sound. In saying *lint*, for example, the tail end of [n] will be devoiced in anticipation of the following voiceless [t]. Figure 4.1 presents a spectrogram that illustrates how the utterance *Weren't you here yesterday?* appears acoustically. There isn't any separation between sounds within a word or between one word and the next. In the same way, the acoustic signal picked up by an infant's ears is continuous, and part of a child's task is to sort out words within sentences and sound segments within a word.

What a Child Must Learn

Children pass through stages in learning words. Some children appear to take up phrases and clauses in utterances as whole units and later to dissect them into parts (a gestalt approach). Others manage a more analytic approach from the start, taking up words directly and constructing phrases and clauses from them as necessary. All children eventually sort utterances into distinct units of meaning.

What kind of information must a child learn about the sounds of a word? What is needed just to recognize a word? For one thing, a child must recognize pronunciations of a given word by different people as the same word, as we indicated above. Even when saying "the same thing," the acoustic signals differ when produced by an older person and a younger one or by a man and a woman—and therefore the acoustic signal heard by the infant also differs significantly. To understand speech it is essential to disregard certain voice characteristics and to disregard the particularities of volume, speed, and pitch of different speakers. For another thing, a child must observe not only the sound segments in a word but the order in which they occur—not only that *cat* contains the three sound segments [k], [æ] and [t] but it has them in that order. And, of course, in some fashion a child must know the phonological features that characterize each segment—for example, that [k] is a voiceless-velar-stop and [t] a voiceless alveolar stop. Perhaps to put it in a way easier to understand, the child must know that [k] and [t] are different sounds, though they are very much alike in most ways.

Phonemes and Allophones

To understand what is being said, a child must disregard differences in the voice quality of different speakers, as we've said. Beyond that, a child must also learn that the same sound is pronounced differently in different contexts—in other words, that physically different sounds may be different pronunciations of "the same sound." We've already mentioned the different /t/ sounds in *tin* and *twin*. Here are some additional examples. Without being aware of it, English speakers pronounce the words *keep* [kip] and *cop* [kɑp] with different [k] sounds. If you alternately pronounce *keep* and *cop*, focusing on where your tongue touches the roof of your mouth at the very beginning, you'll be able to notice that it touches the velum farther back for *cop* than for *keep*. The reason is that the [ɑ] sound of *cop* is a back vowel and the [i] sound of *keep* is a front vowel and, in anticipation of the back vowel to follow, [k] is pronounced farther back for *cop* than when [k] precedes the front vowel in *keep*.

Try It Yourself Prepare to say *keep*. Then, once your tongue is in position for the initial sound of *keep*, say *cop* instead. You'll find that you must reposition your tongue to do it. If you say *cop* from the initial position for *keep*, it will sound peculiar or foreign. The need to reposition the back of your tongue to achieve a natural pronunciation demonstrates that the initial sounds of *keep* and *cop* are not identical.

As another example of a sound being pronounced differently in different environments, note the sounds represented by <p> in *pot* and *spot* or *poke* and *spoke*. Hold the back of your hand up to your mouth when saying these word pairs aloud and you'll feel a difference in the puff of air that accompanies the sounds represented by <p>. In *pot* and *poke* it's aspirated, and we represent that puff of air, the aspiration, by a superscripted *h*: as [p^h]. But after [s], the sound represented by <p> is not aspirated, and we represent it as [p]. Thus, a more detailed phonetic representation of these words would be [p^hɑt] for *pot* and [spɑt] for *spot*.

Try It Yourself The aspiration accompanying the [p^h] sound in *pot* and *poke* is strong enough to blow out a lighted match held in front of the mouth. The [p] sound following [s] in *spot* and *spoke* is not aspirated and will not blow out a match. Try saying *spot*, *spot*, *spot*, followed by a single *pot*. If everything is positioned correctly, saying *spot* will leave the match burning, but saying *pot* will blow it out. Be very careful when doing this!

In discussing [p^{h}] and [p], we've noted that they occur in different positions, or environments, within a word. The list of words below illustrates the environments in which aspirated [p^{h}] and unaspirated [p] occur.

pot	[p^{h}ɑt]
poker	[p^{h}okər]
plate	[p^{h}let]
spot	[spɑt]
spine	[spaɪn]
split	[splɪt]

Aspirated [p^{h}] occurs at the beginning of words (*pot, poker, plate*) but not after [s], where only unaspirated [p] occurs (*spot, spine, split*). When you have two sounds and neither can occur where the other one occurs in a word, we say they occur in **complementary distribution**.

In the words listed above, aspirated [p^{h}] occurs only word initially and unaspirated [p] only after [s]. English doesn't have any words with an initial unaspirated [p] or any with [p^{h}] after [s] in the same syllable. By definition, then, [p^{h}] and [p] are in complementary distribution. And because [p^{h}] and [p] do not occur in the same position in a word, they cannot be the basis for distinguishing one word from another. We say they cannot contrast, are not contrastive. In order for [p^{h}] and [p] to contrast, there would have to be a pair of words like [p^{h}ɑt] and [pɑt] or [sp^{h}ɑt] and [spɑt] with different meanings, but there aren't any such pairs. We conclude that [p^{h}] and [p] cannot signal a contrast between English words. Instead, they are different realizations of the same unit in the English sound system. To different realizations of the same sound unit we give the name *allophones*. Thus [p^{h}] and [p] are allophones of /p/, and we call the unit a *phoneme*: [p^{h}] and [p] are allophones of the phoneme /p/ in English.

Note: To enclose phonemes we have started using slanted lines / / and to enclose allophones we use square brackets [ ]. We will continue this practice, but sometimes we will need to choose one representation or the other when either one would serve. We use angled brackets < > to enclose letters, not sounds.

A **phoneme** is a structural unit in the sound system of a language, and the set of phonemes in a language is the set of distinctive, or contrastive, sounds it exploits. **Allophones** of a phoneme are realizations of a single structural unit. They cannot create different words, so they are noncontrastive. Native speakers of a language perceive the allophones of a phoneme in their language as the same sound despite allophones having physical differences in pronunciation. Thus English speakers regard [p^{h}] and [p] as the same sound unit, namely /p/. Likewise for the two [k] sounds of *cop* and *keep*—they appear in complementary distribution (the first before back vowels and the second before front vowels) and could not serve to differentiate between words: they are therefore allophones of the phoneme /k/ and are perceived by English speakers as the same sound unit.

Besides the aspirated and unaspirated allophones of /p/, there is a third voiceless bilabial stop, which can occur in English words like *map* and *stop*. This allophone can occur at the end of a word at the end of an utterance, as in *Where's* *the map?* or *Don't stop!* In this environment, the lips may remain closed, resulting in a sound with no audible release, and we represent this allophone of /p/ as [p˺]. Now, you may have noted that this fact slightly complicates what we have already said because both unaspirated [p] and unreleased [p˺] can occur word finally: they are *not*, therefore, in complementary distribution. When two phonetically similar sounds such as [p] and [p˺] *can* occur in the same position in a word but do not create contrasting words, they are said to occur in **free variation**. At the end of an utterance, English speakers can pronounce *map* as [mæp] or [mæp˺] and *stop* as [stɑp] or [stɑp˺]. Both unaspirated [p] and unreleased [p˺] are allophones of /p/, and /p/ therefore has three allophones. But all of them are voiceless *bilabial* stops. The same thing is true for aspirated, unaspirated, and unreleased voiceless *alveolar* stops: they are allophones of /t/. Likewise, aspirated, unaspirated, and unreleased *velar* stops are allophones of /k/. To recapitulate, then, the allophones of a phoneme occur in complementary distribution or free variation, and in no case can a change of meaning be signaled by different allophones of the same phoneme.

Distribution of Allophones

It is helpful to view a phoneme as an abstract element in the sound system of a language. Perhaps thinking of a phoneme as a skeletal unit with bones but not a fully fleshed out pronunciation would be helpful. How a phoneme will be realized—fleshed out—depends on its environment. For example, the phoneme /p/ would have the skeletal features *voiceless bilabial stop* and would be aspirated in some environments (*pot, pill*) but unaspirated in others (*sport, spell*) and sometimes unreleased (*flop, blip*). Its realization can be fully specified only when its position in a word or utterance—its environment—is known.

We have just seen that, depending on where in a word it occurs, one allophone of a phoneme or another allophone is required. If you examine the sets of words in Table 4.1, you'll see that the picture is more interesting than that. In these words the accent mark (´) indicates the syllable that bears the primary stress (as in *rídicule* versus *ridículous*).

All the words in Table 4.1 share a feature: they all have a syllable-initial /p/. In column A primary stress falls on the first syllable—and /p/ is aspirated. In column B, /p/ starts a stressed syllable that is not the first syllable of the word—and /p/ is aspirated. Columns A and B thus demonstrate that /p/ is aspirated when it initiates a stressed syllable. In the words of column C, word-initial /p/ is aspirated although primary stress does *not* fall on the first syllable. In sum, /p/ is aspirated when it initiates a stressed syllable or when it is word initial. The words in column D demonstrate that /p/ is not aspirated when it begins an unstressed syllable within, but not at the beginning, of a word. In sum, the English phoneme /p/ is always aspirated word initially (whether in a stressed or unstressed syllable—*pérfect* and *perféction*) and when it initiates a stressed syllable within a word (*compúter* and *Topéka* but not *táper* or *cópious*).

Given these observations, we can be more precise in describing the distribution of the allophones of /p/, taking account of stress patterns. This we do in Table 4.2.

Table 4.1 *Two Allophones of /p/ in English Words*

A	B	C	D
[pʰ]	[pʰ]	[pʰ]	[p]
pédigree	empórium	petúnia	rápid
pérsonal	compúter	patérnal	émpathy
pérsecute	rapídity	península	competítion
pílgrimage	compétitive	pecúliar	computátional

Table 4.2 *Two Allophones of English /p/*

Phoneme	Allophones	Distribution
/p/ →	[pʰ]	In syllable-initial position in a stressed syllable and in word-initial position
/p/ →	[p]	Elsewhere

Try It Yourself Pronounce the words *papaya* and *purpose* aloud to determine which of the four occurrences of /p/ are aspirated and which are not. Using the information about the distribution of /p/ allophones as described in Table 4.2, explain their distribution in these two words.

Contrast the facts about aspirated [pʰ] and unaspirated [p] with the facts about /p/ and /s/. A child aiming to say *sat* but saying *pat* instead would fail to observe one of the significant differences in English pronunciation—because /p/ and /s/ are distinct phonemes and, as such, can distinguish words, as in these pairs:

A	B	C	D
/pɪt/ pit	/pʌn/ pun	/læpt/ lapped	/sip/ seep
/sɪt/ sit	/sʌn/ sun	/læst/ last	/sis/ cease

The words in each pair differ by only a single sound but have altogether different meanings. Two words that differ by only a single sound constitute a **minimal pair**: note that the difference is one in *sounds*, not *spelling*. Minimal pairs are valuable in identifying the contrastive sounds of a language—its phonemes. Each minimal pair above demonstrates that /s/ and /p/ are distinct phonemes of English and not variants (allophones) of a single phoneme. The articulatory descriptions of /s/ and /p/, as you will recall from the previous chapter, show that they differ in place of articulation and manner of articulation but both are voiceless.

	/s/	/p/
Voicing	-voice	-voice
Place of Articulation	alveolar	bilabial
Manner of Articulation	fricative	stop

To take another example, /s/ and /b/ differ from one another in place and manner of articulation as well as in voicing.

	/s/	/b/
Voicing	-voice	+voice
Place of Articulation	alveolar	bilabial
Manner of Articulation	fricative	stop

The fact that /s/ (a voiceless alveolar fricative) and /b/ (a voiced bilabial stop) contrast in English (as in the minimal pairs *sorrow/borrow* and *grass/grab*) demonstrates they are distinct phonemes.

Realizations (allophones) of a single phoneme share most phonetic features but differ in at least one feature—voicing (voiced vs. voiceless), aspiration (aspirated vs. unaspirated), manner of articulation (for example, stop vs. fricative), or place of articulation (for example, dental vs. alveolar). Consider the words *tap* [t^{h}æp] and *tab* [t^{h}æb]. The difference in meaning between them must be signaled by the only difference in sound: the voicing that differentiates [b] from [p]. In fact, the [p]/[b] distinction serves to differentiate hundreds of English word pairs, including *patter/batter* and *rope/robe*. In other sounds of English the voicing feature serves to differentiate thousands more.

The English phonemes /p/ and /b/ contrast in word-initial position (*pit/bit*) and word-final position (*slop/slob*), as we've seen. Sometimes, however, two sounds will contrast in some positions but not all. Two sounds are distinctive—different phonemes—if they contrast in *any* position. The minimal pair *pit/bit* (and many others) demonstrates that /p/ and /b/ are distinct English phonemes. But English does not have any pair of words where /p/ and /b/ contrast in the position following /s/. In other words, there isn't a pair like *spat/sbat* or *spill/sbill*, but that fact does not undercut the constrastiveness of /p/ and /b/ as separate phonemes. (We return to this point later.)

Now, to illustrate how sounds can be organized differently within the sound systems of different languages, let's consider bilabial stops in Korean. Like English, Korean has three bilabial stops—[p^h], [p], and [b]—as in these words:

[p^{h}ul]	'grass'
[pul]	'fire'
[pəp]	'law'
[mubəp]	'lawlessness'

Because [p^{h}ul] and [pul] mean different things and differ in only a single sound segment, they constitute a minimal pair and demonstrate that [p^h] and [p] contrast and therefore belong to different phonemes in Korean. (Recall that in English [p^h] and [p] are allophones of a single phoneme.) Unlike English, Korean does *not* have a minimal pair in which [p] contrasts with [b]—there are no such pairs as *pit/bit* or *mop/mob*. Instead, in Korean [p] and [b] occur in complementary distribution: [b] occurs only between vowels or other voiced segments, as in [mubəp], where [p] never occurs. Given their phonetic similarity, this demonstrates that [p] and [b] are allophones of a single phoneme. (Note in the list above that the morpheme meaning 'law' occurs as [pəp] and [bəp].)

Table 4.3 shows that, between them, the two Korean phonemes /p/ and /p^h/ have three realizations. It represents the fact that [p] and [b] are

Table 4.3 *Three Sounds of Korean*

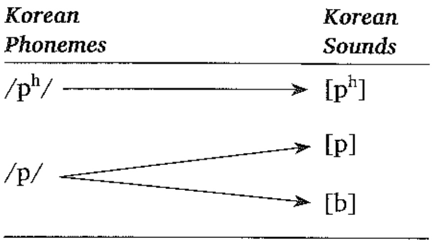


Table 4.4 *Three Sounds of English*

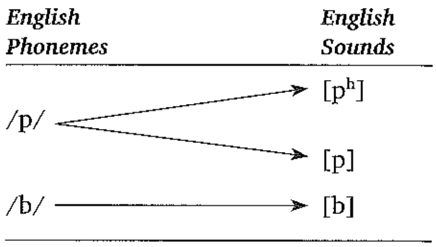
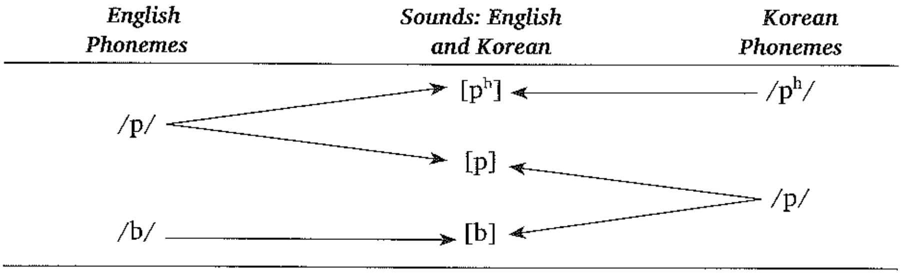


Table 4.5 *Three Sounds of English and Korean Compared*



realizations—allophones—of /p/ and that [pʰ] is a realization—an allophone—of /pʰ/ (whereas [p] is not). Table 4.4 shows that, between them, the two English phonemes /p/ and /b/ have three realizations. It represents the fact that [p] and [pʰ] are allophones of /p/ and that [b] is an allophone of /b/. Those patterns illustrate that, among these sounds at least, voicing is not contrastive in Korean but aspiration is contrastive. In English, on the other hand, voicing is contrastive but aspiration isn't.

Table 4.5 combines the facts represented in Table 4.3 and Table 4.4. It captures the fact that English and Korean both have the three sounds [p], [pʰ], and [b]. But they function quite differently in the sound systems of those languages. In Korean, [p] and [b] function as the same sound unit, and [pʰ] is a different one; in English, [p] and [pʰ] function as the same sound, while [b] is a different one.

Phonological Rules and Their Structure

You may be aware that French has nasal vowels. In fact, the nasal vowels of French are so well known that you may think English lacks nasal vowels. In fact, English does have nasal vowels, as in the words of column B in Table 4.6. The vowels in those words are pronounced with air passing through the nose, as well as through the mouth, whereas the words in column A have oral vowels (pronounced through the mouth). When you pronounce the words of column B, air from the lungs exits through the nasal passage; when that passage is blocked, the sound of the vowel changes perceptibly.

Try It Yourself Pinch your nose closed while saying the words in Table 4.6. You'll discover that for the words in column A it will make no perceptible difference, but for those in column B it will make a striking difference.

Table 4.6 *Oral and Nasal Vowels in English*

A	B
sit	sin
pet	pen
light	lime
brute	broom
sitter	singer

If you search out nasal vowels in English words, you'll discover that *all* of them precede one of the nasal consonants /m/, /n/, or /ŋ/. That's another way of saying that the distribution of nasal vowels in English is rule-governed and predictable: English vowels are nasalized before a nasal consonant. Since the distribution is rule-governed, the occurrence of nasal vowels cannot signal a meaning distinction. There is no pair of English words distinguished solely by virtue of one having a nasalized vowel and the other having the same vowel not nasalized. By contrast, in French and some other languages, nasalization can signal a difference in meaning because its distribution is not predictable by rule. Two phonetically similar sounds whose distribution with respect to one another is predictable by rule constitute allophones of a single phoneme.

Phonological rules (like the one governing nasalization in English) have this general form:

A → B / C___D

You can read this rule as "A is realized as B in the environment following C and preceding D" or, more simply, "A becomes B following C and preceding D." The dash between C and D on the right side of the arrow indicates where what is on the left side of the arrow would occur. A, B, C, and D can be specified informally or in terms of phonological features. In cases where it is unnecessary to specify both C and D, one of them will be missing. For example, the phonological process of nasalization in English can be represented as follows:

Nasalization Rule
vowel → nasal /___ nasal
(In words: vowels are nasalized when they precede nasals.)

As we said earlier, in acquiring a word a child must learn which phonemes it contains and the order in which they occur. As the English *cop/keep* alternation shows for the allophones of /k/, and as the *poke/spoke* alternation shows for the allophones of /p/, a child must also learn the rules that determine the particular allophones of a phoneme depending on its position in a word and the character of nearby phonological features such as nasal or voicing or position in a word or syllable. For example, for each word in Table 4.6, a child doesn't have to learn whether its vowel is nasalized or not: the nasalization rule will apply only to the words in Column B because only those words meet the structural conditions specified by the nasalization rule.

The situation for a child acquiring Korean [p] and [b] is parallel to that of an English-speaking child acquiring nasal vowels. Since [p] and [b] never contrast in Korean, they are allophones of a single phoneme and have only a single representation in the lexicon. Of course, speakers must also know the phonological rule that specifies the distribution of the allophones: [b] between vowels (and other voiced sounds) and [p] elsewhere. The alternative to having a single representation in the lexicon for [p] and [b] would require a specific differentiation between these sounds in every word that contains either of them. For example, */pəp/* 'law' and */mubəp/* 'lawlessness' would have different specifications for [p] and [b]. But to have different specifications for [p] and [b] in the mental lexicon of a Korean speaker would be equivalent to an English speaker's having different specifications for the different /k/ sounds of *cop* and *keep*, the different /p/ sounds of *poke* and *spoke*, and the different /i/ sounds of *seat* and *seen*. Instead, each phoneme in a language is represented in the lexicon by a single form (called an underlying form, to be discussed shortly). Native speakers internalize the phonological rules specifying the distribution of allophones and automatically apply them wherever the conditions for rule application are satisfied.

Generalizing Phonological Rules

Until now we have considered phonological rules as though they were formulated to apply to particular sound segments. But they are more general than that (and actually easier to learn, then). Let's revisit aspiration as it accompanies the realization of /p/ in English words like *poke* and *oppose*:

1. Aspiration Rule for /p/:
   -voice
   bilabial → aspirated / word initially and initially in stressed syllables
   stop

This rule says that a voiceless bilabial stop is aspirated in specific environments. In *poke* it would trigger aspiration because /p/ occurs word initially, and in *oppose* it would trigger aspiration because /p/ initiates a stressed syllable.

Why a Foreign Accent? Part 1

Consider a native speaker of English who knows no French and has been introduced by a French speaker to neighbors named Pierre and Paulette. French speakers do not aspirate initial voiceless stops so the French speaker will pronounce the /p/ in those names without aspiration. Despite the fact that the English speaker has *not* heard aspiration, he or she will tend to pronounce *Pierre* and *Paulette* with an aspirated [p^h]—in violation of the phonological rules of French but conformity to English rules. Subconsciously and automatically applying the phonological rules of one's native language to words in another language contributes significantly to speaking with a foreign accent and identifying one as nonnative.

On the flip side, when speaking a foreign language you may fail to make a necessary distinction. English speakers don't have to learn separately for *cop* and *keep* which *k* sound to use because English distinguishes these sounds by rule. But in other languages these two sounds may be distinct phonemes and serve to contrast words. The IPA symbol for the initial sound of *cop* is [k] and of *keep* is [c]. In Basque, Malay, and Vietnamese, among others, it is critical to know which velar stop occurs in a word, just as English speakers must know whether a word contains /p/ or /t/: in those languages, [k] and [c] are not distributed by rule and are therefore not predictable. They must be learned for individual words.

If, besides words like *poke* and *oppose*, you examine English words with other stop consonants, you'll discover that /t/ and /k/ are aspirated in precisely the same environments in which /p/ is aspirated. Since /p, t, k/ have parallel distributions of aspirated allophones, English would appear to need two additional rules as parallels to the Aspiration Rule for /p/: one for /t/ and another for /k/.

2. Aspiration Rule for /t/:
   -voice
   alveolar → aspirated / word initially and initially in stressed syllables
   stop

3. Aspiration Rule for /k/:
   -voice
   velar → aspirated / word initially and initially in stressed syllables
   stop

But in fact English doesn't need three parallel rules because application of these rules exhausts the list of voiceless stops in English—there are no others besides /p, t, k/. That allows the three processes of aspiration to be captured in a single, more general rule covering /p, t, k/:

English Aspiration Rule
-voice
→ aspirated / word initially and initially in stressed syllables
stop
(Voiceless stops are aspirated in word-initial position and initially in a stressed syllable.)

The English aspiration rule omits place of articulation (bilabial, alveolar, and velar) but maintains the voiceless and stop features. Because the English aspiration rule specifies voiceless stop but not place of articulation, it will apply to all voiceless stops irrespective of their place of articulation. It is thus a more general rule.

The more general a rule, the simpler it is to state using feature notation. Using the features voiceless bilabial stop (instead of /p/) seemed initially to complicate the statement of the Aspiration Rule for /p/. But using features also allowed us eventually to recognize that the aspiration rules for /t/ and /k/ were essentially the same as for /p/ and to collapse all three rules into a single—and simpler—one than any of the individual rules for /p/, /t/, or /k/. For this reason and others, linguists regard internalized phonological rules as being specified not in terms of phonemes such as /p/ or /t/ or in terms of allophones such as [p] and $[p^h]$ but in terms of feature sets such as [-voice] and [stop].

Natural Classes of Sounds

As we have just seen for aspiration in English, phonological processes do not operate on individual sounds but on sets of sounds. English doesn't have three aspiration rules—one each for /p/, /t/, and /k/—but a single aspiration rule that covers all voiceless stops. A set of sounds such as voiceless stops that can be affected by a single phonological rule is called a **natural class** and can be characterized using fewer features than would be needed to describe any of its individual members. For example, /p/ has the features [-voice], [bilabial], and [stop], while the larger set /p, t, k/ can be characterized solely by the features [-voice] and [stop]. Because all places of articulation for voiceless stops are included in the natural class, place of articulation does not need to be specified. A natural class is the set of all sounds sharing a set of features—all and only those sounds. The sounds /p, t, k/ are voiceless stops, sharing the features [-voice] and [stop], and no other sounds in English have only those features. Thus, /p, t, k/ constitutes the natural class of voiceless stops in English.

Now consider /p, t, k, b, d, g/. All the sounds in that set share the feature [stop], and there are no other stops in English. So /p, t, k, b, d, g/ constitutes the natural class of English stops. By contrast, the set of sounds /p, t, k, b, d/ does not constitute a natural class because, while those sounds do share the feature [stop], so does /g/, which isn't included in the set. Similarly, the sounds /p, t, k, z/ do not constitute a natural class because [-voice] and [stop] accurately name features of /p, t, k/, but /z/ is neither voiceless nor a stop, and any additional feature needed to characterize /z/ would also characterize other sounds. Notice, too, that to characterize a set such as /p, t, k, m, n, ŋ/, which includes voiceless stops and nasals, would require an "either/or" description: *either* [voiceless] [stop] *or* [nasal]. No combination of features uniquely specifies just the six sounds /p, t, k, m, n, ŋ/ so that set does not constitute a natural class. Because phonological processes apply to natural classes of sounds, we would not expect to find a rule in any language that applies to, say, /p, t, k, b, d/ or /p, t, k, m, n, ŋ/ or any other set of sounds that does not constitute a natural class.

Underlying Forms

Thanks to internalized rules that yield the correct realizations—that is, allophones—for the phonemes in a given word, children eventually produce entries in their lexicons that contain the equivalent information of the representations in Table 4.7. Such representations are called **underlying forms**, and we write them between slanted lines, using the notation for phonemes. A **surface form** or surface realization results from application of phonological rules to the underlying form, and the surface form underlies a word's pronunciation. In some examples in Table 4.7, the surface form is the same as the underlying form because no phonological rule (of those discussed in this chapter) is applicable, and we have indicated "none" in the Rule column.

Table 4.7 *Underlying and Surface Forms for Six English Words*

Underlying Form	Rule	Surface Form	Written Form
/kʌlər/	aspiration	[kʰʌlər]	color
/bʊk/	none	[bʊk]	book
/bit/	none	[bit]	beet
/ʌp/	none	[ʌp]	up
/spɪn/	nasalization	[spɪ̃n]	spin
/pɪn/	aspiration, nasalization	[pʰɪ̃n]	pin

Rule Ordering

One additional rule will illustrate a point about the organization of phonological rules in the internal grammar. Consider the following words:

A	B	A	B
glop	glob	mop	mob
write	ride	treat	treed
rope	robe	clout	cloud
tap	tab	root	rude
lock	log	sack	sag
flack	flag	clock	clog

If you listen attentively while pronouncing these words aloud, you'll notice that the vowels in column B are slightly longer in duration than those in the corresponding words of column A. The IPA represents long vowels by placing a specific kind of colon after them, as in [uː]. Because English has no minimal pair such as [piːt] and [pit] or [bæt] and [bæːt], vowel length does not contrast in English. Instead, it can be specified by rule and is therefore predictable. Looking past spelling, you'll notice that all the words of column A end with a voiceless consonant /p/, /t/, or /k/, and all those in column B end in a voiced consonant /b/, /d/, or /g/. English vowels have longer duration when they precede voiced consonants. Using V to represent vowel, we can state the rule as follows:

Vowel Lengthening Rule (initial approximation)

V → Vː /___ stop
+voice

(Vowels are lengthened preceding voiced stops.)

As a consequence of this rule, the following processes take place in English:

æ → æː	/ ___ /b/	(as in *tab* versus *tap*)
ɛ → ɛː	/ ___ /d/	(as in *bed* versus *bet*)
o → oː	/ ___ /g/	(as in *code* versus *coat*)
aɪ → aːɪ	/ ___ /d/	(as in *slide* versus *slight*)

(Note that this rule of English applies to all vowels—including diphthongs like /aɪ/.)

Try It Yourself You may suspect that if a following voiced *stop* lengthens a vowel in English, a following voiced fricative or affricate might have the same effect. Find out by saying these word pairs aloud: *proof/prove*; *ether/either*; *Bruce/bruise*; *fishin'/fission*; *batch/badge*. Does that list illustrate all voiced/voiceless pairs of fricatives and affricates in English? What did you discover about whether vowels are lengthened before voiced fricatives and affricates?

If, besides stops, you examined fricatives and affricates in word pairs such as those given in the Try It Yourself on the previous page, you'd discover that English vowels are lengthened preceding stops, fricatives, and affricates. In fact, they are lengthened before *all* voiced consonants. That means that the vowel lengthening rule can be generalized as follows, using V to represent vowel and C to represent consonant:

English Vowel Lengthening Rule

V → Vː / ___ C
+voice

(Vowels are lengthened preceding voiced consonants.)

Because vowel length in English can be specified by rule, it is predictable and need not be learned individually for each word.

On the other hand, vowel length is *not* predictable in all languages. It must be learned word by word. For example, Fijian has minimal pairs distinguished only by vowel length. Compare the members of these pairs (in which long vowels are represented by doubling): *oya* meaning 'he, she,' *oyaa* meaning 'that (thing)'; *dredre* 'to laugh,' *dreedree* 'difficult'; *vakariri* 'to boil,' *vakaririi* 'speedily.' Because vowel length in Fijian words cannot be assigned by rule, it is contrastive—that is, phonemic—in that language.

Now consider the following word pairs, paying attention to how the column A pronunciations differ from the corresponding ones in column B. At least for most speakers of American English, the pronunciation difference is not the one suggested by the spellings <t> and <d>; instead, the difference is in the length of their vowels. In some varieties of American English, the first vowel in each word of column B is longer than the corresponding vowel of column A.

A	B
writer	rider
bleater	bleeder
rooter	ruder

The medial consonants <t> and <d> in the words above do not represent different pronunciations because in certain environments words with /t/ and /d/ between vowels Americans tend to use a tap allophone (see Chapter 3). In the pronunciation of medial /t/ and /d/, the tip of the tongue rapidly taps the alveolar ridge. (Go on. Try it.) Because the tap realizations of /t/ and /d/ are identical to one another (and are represented by IPA [ɾ]), the difference of pronunciation that might have resulted from the underlying distinction between /t/ and /d/ is **neutralized**. The distinction is maintained in *write* [raɪt] and *ride* [raːɪd], *bleat* [blit] and *bleed* [bliːd], and *root* [rut] and *rude* [ruːd] but neutralized when underlying /t/ and /d/ meet the conditions of the tapping rule (as in the examples above). Using V to represent any vowel, the tapping rule for American English can be represented as here:

Tapping Rule

alveolar stop → tap / V ___ V
unstressed

(/t/, /d/ are realized as [ɾ] between two vowels, the second of which is unstressed.)

Interestingly, despite the fact that the tapping rule neutralizes the distinction between /t/ and /d/ in this environment, some American dialects do pronounce the words of column A and column B differently. By applying the tapping rule and the lengthening rule, some speakers of American English pronounce the words in column B with a vowel of longer duration even though there is no difference in the pronunciation of the tap consonant. Here's an explanation for how it may work.

We've described tapping and lengthening rules, and you understand that they may operate on the same word. Let's examine how the rules interact to produce a surface form. Assume that the underlying forms in the lexicon are /raɪtər/ for *writer* and /raɪdər/ for *rider*. We can represent the derivation of the surface forms as in Table 4.8. (When the form of a word does not meet the requirements specified in a rule, that rule is not applicable, so we write NA.) From the underlying forms and the application of the rules in the order shown (first lengthening, then tapping), the surface forms [raɪɾər] and [raːɪɾər] are produced, and these are the correct realizations for the words among some speakers. Call them speakers of dialect A.

If the same rules are applied in the reverse order (first tapping, then lengthening), the surface forms will be different. Here's how it works: because the tap sound is voiced, the vowel preceding it is lengthened in both words. As Table 4.9 shows, this is precisely what happens for speakers of what we can call dialect B.

Table 4.8 *Derivation of* Writer *and* Rider *in Dialect A*

	Writer	**Rider**	
Underlying form	/raɪtər/	/raɪdər/	(input)
Lengthening rule	NA	applies ↓	
Derived form	[raɪtər]	[raːɪdər]	(output/input)
Tapping rule	applies ↓	applies ↓	
Surface form	[raɪɾər]	[raːɪɾər]	(output)

Table 4.9 *Derivation of* Writer *and* Rider *in Dialect B*

	Writer	**Rider**	
Underlying form	/raɪtər/	/raɪdər/	(input)
Tapping rule	applies ↓	applies ↓	
Derived form	[raɪɾər]	[raɪɾər]	(output/input)
Lengthening rule	applies ↓	applies ↓	
Surface form	[raːɪɾər]	[raːɪɾər]	(output)

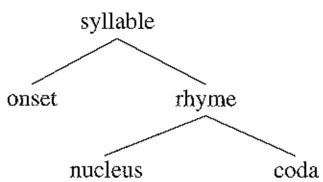
The identical surface forms [raːɾər] 'writer' and [raːɾər] 'rider' that are derived by applying the tapping rule prior to the lengthening rule would not be correct for dialect A. In dialect A, *writer* and *rider* are not pronounced alike. Instead, *rider* has a longer vowel than *writer*. Beginning with the same underlying forms and the same two rules, different speakers produce different surface forms because they apply the rules in different orders. Evidence such as this has led some linguists to theorize that rule ordering (or an equivalent process in the brain) is part of the organization of phonological rules.

Note that the forms resulting from the derivation in Table 4.9, though incorrect in dialect A, are correct in dialect B. This illustrates how speakers of different dialects could share underlying forms and phonological rules but produce different surface forms and pronounce words differently as a result of applying the phonological rules in different orders. Dialects that applied the lengthening rule before the tapping rule would exhibit different forms of *writer* and *rider* (Table 4.8). Dialects that applied tapping before lengthening would exhibit identical forms, both with a long vowel (Table 4.9).

Syllables and Syllable Structure

So far we've said little about how sounds are organized within words although our analyses have in fact presumed a certain organization, as you'll see. We've relied on the obvious fact that segments in a word occur as a sequence *abcdef*, but that isn't the whole story. When we analyzed aspiration, we also relied on the fact that words consist of one or more syllables. Now let's take a closer look at the structure of syllables: we'll discover that within a syllable sounds are organized both sequentially and hierarchically.

Syllable isn't a tough notion to grasp intuitively, and there is considerable agreement in counting syllables within words. Probably most readers of this book would agree that *trout* has one syllable, *salmon* two, *halibut* three, and *barracuda* four. (How many syllables do you hear in the word *barracudina*?) Despite the relative ease of counting syllables intuitively, technical definitions can be challenging. Still, there is agreement that a **syllable** is a phonological unit consisting of one or more sounds and that syllables are divided into two parts—an onset and a rhyme (sometimes spelled <rime> in this context). The **rhyme** consists of a peak, or nucleus, and any consonants following it. The **nucleus** is typically a vowel although the class of consonants called sonorants may function as the nucleus of a syllable. **Sonorants** include nasals (/m, n, ŋ/) and liquids (/r, l/). Within a syllable, any consonants that precede the rhyme constitute the **onset**, and any consonants that follow the nucleus as part of the rhyme constitute the **coda**.



The tree to the left represents the structure of a syllable as just described. In some other books, and commonly in technical discussions, the word syllable in the tree is replaced by σ, the Greek letter sigma.

Not all syllables have an onset, and not all rhymes have a coda. That means the only essential element of a syllable is the nucleus. Because a single sound segment can constitute a syllable and a single syllable can constitute a word, words may consist of a single vowel—but you already knew that from knowing the English words *a* and *I*.

Try It Yourself In the lines below from a Shakespearean sonnet, determine whether the rhymed parts of the rhymed words meet the definition of rhyme formulated on the previous page. In other words, do these Shakespearean rhymes consist solely of the nucleus and coda of a syllable? *Sometime too hot the eye of heaven shines, / And often is his gold complexion dimm'd; / And every fair from fair sometime declines, / By chance or nature's changing course untrimm'd.*

Table 4.10 gives some English words with one, two, three, and four syllables. (We use a period to separate syllables within a word.)

Using their ordinary spellings, the trees in Figure 4.2 illustrate the linear and hierarchical syllable structure of the words *fat*, *even*, and *loveliest*. As an exercise, you are encouraged to draw the tree that represents the word *respectively*.

Sequence Constraints

Notice in the phrase below that English syllables exhibit several patterns of consonants (C) and vowels (V).

in	a	pre.vi.ous	cap.tion
ɪn	ə	pri.vi.əs	kæp.ʃən
VC	V	CCV.CV.VC	CVC.CVC

In the expression *in a previous caption*, you can see that English permits VC, V, CCV, CV, and CVC syllable types. Still other permissible types can be noted in monosyllabic words like these:

past	/pæst/	CVCC	queen	/kwin/	CCVC
turned	/tərnd/	CVCCC	squirts	/skwərts/	CCCVCCC

Table 4.10 *English Words Divided into Syllables*

1 Syllable	2 Syllables	3 Syllables	4 Syllables
fat	even	loveliest	respectively
[fæt]	[i.vən]	[lʌv.li.əst]	[ri.spɛk.təv.li]
spin	although	potato	accumulate
[spɪn]	[ɔl.ðo]	[p^{h}ə.t^{h}e.ɾo]	[ə.k^{h}jum.jə.let]
through	consists	computer	algebraic
[θru]	[k^{h}ən.sɪsts]	[k^{h}əm.p^{h}ju.ɾər]	[æl.ʤə.bre.ək]

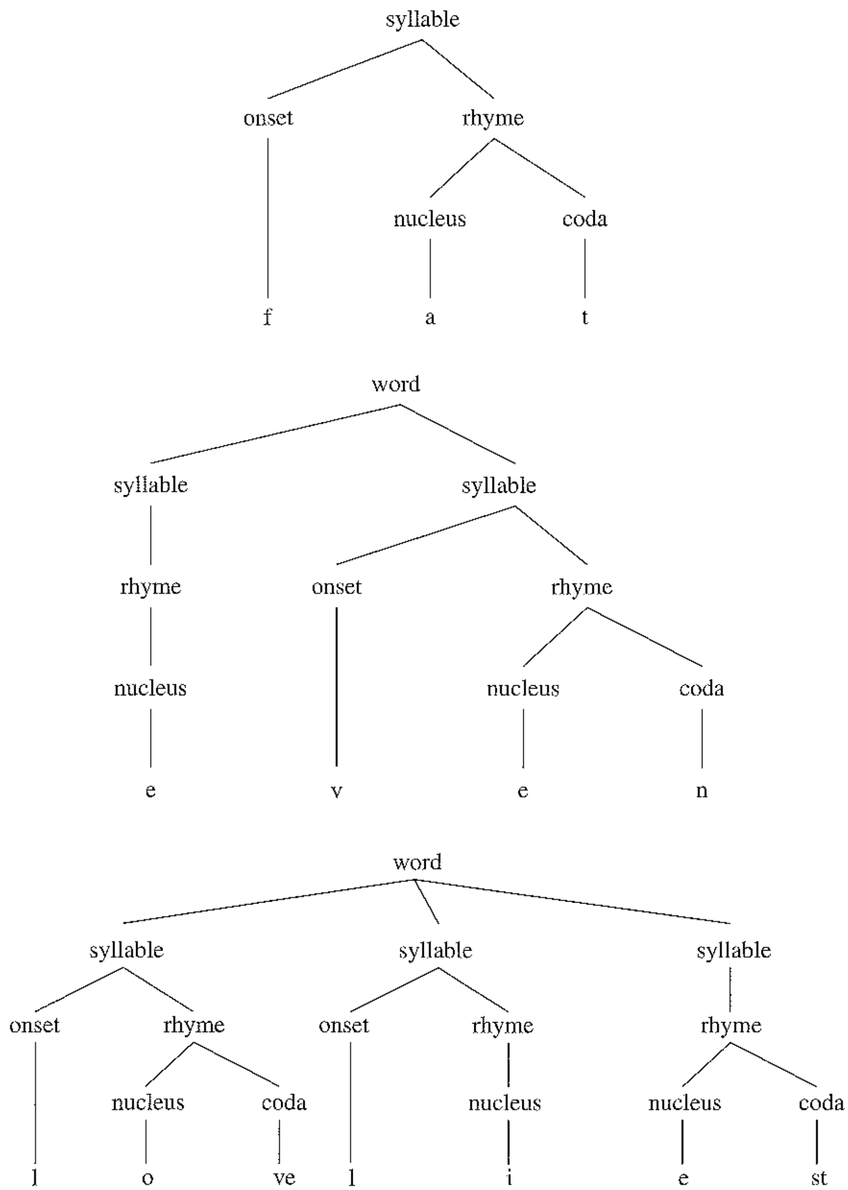


Figure 4.2 Linear and Hierarchical Syllable Structure in *Fat, Even, Loveliest*

Inside a multisyllabic word (for example, at the end of a syllable that is not the last syllable in the word), fewer coda types are permitted.

The sequences of sounds permitted in a syllable differ from language to language, and not every language allows the same variety of syllable types as English allows. Among the world's languages the preferred syllable type is a single consonant followed by a single vowel (CV), in other words a single-consonant onset and a nucleus. Other common types across languages are CVC and V.

(All three types occur in the illustrative phrase *in a previous caption*, given previously.) Polynesian languages like Samoan, Tahitian, and Hawaiian have only CV and V syllable types. Japanese allows CV and V types and also permits CVC if the coda (the second C) is a nasal. Korean permits V, CV, and CVC syllables. Mandarin permits V and CV syllables and, if the coda is [n] or [ŋ], also permits CVC.

Two-consonant onset clusters like those English has in *try*, *twin*, and *stop* are not common in the languages of the world, and it is very uncommon to have onset clusters of more than two consonants as in the CCC pattern represented in English *scream*, *splint*, and *stress*. English speakers can't combine just any two consonants into an onset either. English has a limited range of consonants that can occur as C_1 and C_2 in an onset cluster (C_1C_2) and a narrower range in each position of a three-consonant onset cluster. It is not a coincidence that *scream*, *splint*, and *stress*—all three of our illustrative three-syllable onset clusters—begin with /s/. On top of that, the constraints on three-consonant onsets in English differ from the constraints on codas. Notice, for example, that the word *squirts* /skwərts/ has a three-consonant cluster as the onset and another as the coda. But the coda /rts/ is not a permissible English onset, and /skw/ is not a permissible coda: there could not be English words <rtsoskw> or <rtsuskw>.

Try It Yourself Cite three English words other than *scream*, *splint*, and *stress* that have three consonants in the onset. What sound or sounds do they begin with? Which sounds occur as C_2? What about C_3? Can you think of any onset clusters that have a different C_2 or C_3? What are they?

Try It Yourself The following forms violate the sequence constraints of English and could not serve as English words: *ptlin*, *brkop*, *tsmot*, *ngam*. Add three more impossible words to the list.

The rules that characterize permissible syllable structures in a language are called **sequence constraints** or **phonotactic constraints**, and they determine what constitutes a possible syllable. As a result of such constraints, there are—in addition to the words that exist in a language—thousands more that could exist but happen not to exist, and countless others that *could not* exist because they would violate the sequence constraints of that language. For example, because they would violate the sequence constraints of Hawaiian and of Japanese, the following would be impossible words in either of those languages: *pat* (CVC), *pleat* (CCVC), and *spa* (CCV).

Stress

An aphorism among American linguists points out that "Not every white house is the White House, and not every black bird is a blackbird." The point is that *stress* patterns can be significant. In pronouncing the phrase *every white house*, relatively strong stress is given to both *white* and *house*: *whíte hóuse*. In referring to the official residence of the U.S. president, relatively strong stress is assigned to *White* but only secondary stress to *House*: *Whíte Hòuse*. The stress pattern

Why a Foreign Accent? Part 2

Learning a foreign language whose syllable structure differs from that of our native tongue presents challenges because we tend to impose our native sequence constraints onto pronunciations of words in the foreign language. For example, the onset clusters /st/ and /sp/ are not permitted in Spanish or Persian, and that makes it a challenge for native speakers of Spanish and Persian to pronounce English words like *study* and *speech* as an English speaker would. Instead of /stʌdi/ and /spitʃ/, speakers of Spanish and Persian are often heard to say /ɛs-tʌdi/ and /ɛs-pitʃ/, thereby conforming to the sequence constraints of their native languages. As we noted in Chapter 2, the English words *strike* and *baseball* have been borrowed into Japanese, and in this chapter we saw that Japanese has a basic CV sequence constraint so it isn't surprising that speakers of Japanese pronounce these borrowed words as *sutoraiku* and *beesubooru*, bringing them into conformity with the CV constraint of Japanese: *su.tu.rai.ku* and *bee.su.boo.ru*.

assigned to the name of the president's residence matches that in the word *téachèr*: *Whíte Hòuse*. The stress pattern of the same words in the phrase (*every*) *whíte hóuse* does not. From the fact that stress can vary and that the meanings of the two expressions differ, it follows that stress can be contrastive in English. Below is a list of several other word pairs. The pairs in column A are distinct words—they constitute noun phrases, comprising an adjective and a noun (as well as an article); the stress patterns of the pairs in column B match the pattern of *téachèr*—they constitute compound nouns.

A	B
a bláck bóard	a bláckbòard
a blúe bírd	a blúebìrd
a hígh cháir	a híghchàir
a réd néck	a rédnèck
a jét pláne	a jétstrèam
An íced téa	an íce crèam
a yéllow jácket (clothing)	a yéllow jàcket (a kind of wasp)

English has variable stress, not fixed stress, and so do some other languages, including German. Many other languages have fixed stress, with stress assigned regularly to a particular location in words. In Polish and Swahili, stress typically falls on a word's next-to-last syllable (called the penultimate syllable). Czech words carry stress on the first syllable. French words usually carry stress on the last syllable.

Try It Yourself Identify three additional examples of phrase and compound noun pairs like *blúe bírd* and *blúebìrd* in the preceding columns.

Syllables and Stress in Phonological Processes

We saw above that certain phonological rules depend for their formulation on the syllable (aspiration), on stress (tapping), or on both. English aspiration of the voiceless stops /p, t, k/ occurs "word initially and initially in stressed syllables." Such a formulation assumes that in the brain words are organized by syllable. In turn, that means children must have some grasp of how words are organized into syllables. The tapping rule that produces [raɪɾər] for *writer* and [mɛɾəl] for *metal* also relies on stress, and by now you can probably imagine that the tapping rule could be formulated in terms of syllables instead of vowel segments, which is how we formulated it in the section on "Rule Ordering" above. Current models of word structure use multiple tiers to accommodate phonologically significant levels, including segments, syllables, and stress, but analyses of that kind lie beyond the scope of this book.

Morphology and Phonology Interaction: Allomorphy

Before leaving the subject of phonology, let's return to the pronunciation of the most productive inflectional suffixes of English, which we briefly introduced at the end of Chapter 2. Some striking regularities in the patterns are worthy of further exploration.

English Plural, Possessive, and Third-Person Singular Morphemes

Regular nouns exhibit several realizations of the plural morpheme, as in *lips* [lɪps], *seeds* [sidz], and *fuses* [fjuzəz]. The surface forms for these realizations of the plural morpheme are called its **allomorphs**. As the following three lists demonstrate, the allomorphs of the plural morpheme are determined by the character of the final sound of the singular stem to which it is attached.

Allomorphs of the English 'Plural' Morpheme

[əz]	[s]	[z]
bus-es	tip-s	tab-s
fus-es	cat-s	seed-s
bush-es	book-s	fig-s
peach-es	whiff-s	car-s
judg-es	birth-s	ray-s

These lists indicate the distributional pattern of the plural realizations in English:

1. [əz] after nouns ending in /s, z, ʃ, ʒ, ʧ, ʤ/ (a natural class called *sibilants*)
2. [s] following all other *voiceless* segments
3. [z] following all other *voiced* segments

We will assume that the underlying form of the plural morpheme is /z/. (There are good reasons for making that assumption, but we need not go into them here.) From this underlying form, all three allomorphs must be derivable by general rules that apply to all regular nouns. From an underlying /z/, a rule such as the following would derive the [əz] allomorph that follows sibilants. (*Note*: + marks a morpheme boundary and # marks a word boundary.)

Schwa Insertion Rule A
/z/ → [əz] / sibilant + __#
(Schwa is inserted before a word-final /z/ that follows a morpheme ending in a sibilant.)

In order to derive the allomorph [s] from the underlying /z/ after voiceless sounds, a rule would be needed that changes the voiced /z/ to [s], reflecting the voiceless final consonant sound of the stem. The process of a sound becoming more like an adjacent sound is called **assimilation**, and because underlying /z/ is realized as voiceless following a voiceless sound, we can regard this as a process of assimilation.

Assimilation Rule A
/z/ → -voice / -voice + __#
(Word-final /z/ is devoiced following a morpheme that ends in a voiceless sound.)

In order to derive the correct forms of all regular plural nouns, the schwa insertion and assimilation rules must have considerable generality, and Table 4.11 illustrates the application of these rules in the nouns *coops*, *pieces*, and *weeds*. (NA means a rule is not applicable because a condition necessary for it to apply is missing; slanted lines / / represent underlying forms; square brackets [ ] represent forms derived by application of one or more phonological rules.)

Table 4.11 *Derivation of English Plural Nouns*

	Coops	**Pieces**	**Weeds**
Underlying forms	/kup + z/	/pis + z/	/wid + z/
Schwa insertion	NA	applies ↓	NA
Derived form	[kup + z]	[pis + əz]	[wid + z]
Assimilation	applies ↓	NA	NA
Surface form	[kups]	[pisəz]	[widz]

These rules for deriving the plural forms of regular nouns have even wider applicability. It turns out that for two other extremely common inflectional morphemes of English—the possessive marker on nouns (*judge's*, *cat's*, *dog's*) and the third-person singular marker on verbs (*teaches*, *laughs*, *swims*)—the distribution of allomorphs parallels the distribution of plural allomorphs.

Possessive Morpheme on Nouns

[s]	for *ship, cat, Jack, puff, breath* . . .
[z]	for *crab, pad, bag, arm, John, thing* . . .
[əz]	for *church, judge, fish* . . .

Third-Person Singular Morpheme on Verbs

[s]	for *laugh, leap, eat, kick* . . .
[z]	for *swim, hurry, lean, crave, see* . . .
[əz]	for *teach, tease, judge, buzz, rush* . . .

We posited /z/ as the underlying form of the plural morpheme, and if we posit /z/ as the underlying phonological form of the possessive morpheme and the third-person singular morpheme, the very same rules that derive the correct allomorphic realizations of the plural morpheme will derive the correct realizations of the possessive morpheme on nouns and of the third-person singular morpheme on verbs. (Unlike plurals, some of which are irregular, all nouns have regular possessive morpheme realizations, and all verbs—except *is, has, says*, and *does*—are regular with respect to the third-person singular morpheme.)

English Past-Tense Morpheme

Going beyond the plural, possessive, and third-person singular morphemes just examined, we see a further parallel for another extremely common English morpheme. Pronouncing the verbs below will demonstrate that the inflectional morpheme that marks the past tense of regular verbs in English has three realizations:

[t]	for *wish, kiss, talk, strip, preach* . . .
[d]	for *wave, bathe, play, lie, stir, tease, roam, ruin* . . .
[əd]	for *want, wait, hoot, plant, wade, need* . . .

If we posit /d/ as the underlying phonological form of the past-tense morpheme, only two rules are needed to derive the past-tense realizations of *all* regular verbs.

Schwa Insertion Rule B
/d/ → [əd] / alveolar stop + __#
(Schwa is inserted preceding a word-final /d/ that follows a morpheme ending in an alveolar stop.)

Table 4.12 *Derivation of English Past-Tense Verbs*

	Waved	Wished	Wanted
Underlying form	/wev + d/	/wɪʃ + d/	/wɑnt + d/
Schwa insertion	NA	NA	applies ↓
Derived form	[wev + d]	[wɪʃ + d]	[wɑnt + əd]
Assimilation	NA	applies ↓	NA
Surface form	[wevd]	[wɪʃt]	[wɑntəd]

Assimilation Rule B
/d/ → -voice / -voice + __#
(Word-final /d/ is devoiced following a morpheme that ends in a voiceless segment.)

Notice that, just like the rules for deriving plurals, possessives, and third-person singular allomorphs, the rules for deriving past-tense allomorphs involve schwa insertion and assimilation. Table 4.12 provides derivations of the past-tense forms of the verbs *wave*, *wish*, and *want* as examples.

Underlying Phonological Form of Morphemes in the Lexicon

A word's phonological shape in the lexicon is called its *underlying form*, and generally speaking it is not identical to the surface or realized form, as we've seen. This section explores the phonological form of words as they are thought to exist in a speaker's mental lexicon.

Consonants The same kinds of phonological processes that operate between a stem and an inflectional suffix (for example, between /trɪp/ and /d/ to produce [tʰrɪpt] 'tripped') also operate between a stem and a derivational morpheme. Think about a child who knows the words *metal* and *medal*. In North American English, the sound that occurs in the middle of both words is not [t] or [d] but [ɾ] (an alveolar tap, created when the tip of the tongue taps quickly against the alveolar ridge as in *pity*, *la<u>t</u>er*, *bla<u>dd</u>er*).

A child hearing *metal* and *medal* would have entered exactly what he or she heard into the lexicon—for a speaker of American English that would be /mɛɾəl/ (with a tap) for both words. But after the child had heard someone say a new car was painted *metallic* [mətʰæləl] *red* and recognized that *metallic* contains the morpheme METAL and the derivational suffix -IC (as in *atomic* and *acidic)*, then the alternate pronunciations of the METAL morpheme as [mɛɾəl] and [mətʰæl] would have to be reconciled. One task of a language learner is to posit an underlying form that will efficiently yield the right surface forms as output from general phonological rules.

Next consider what happens when the child subsequently hears someone report that the metallic red car's *medallion* is missing from the hood. For *medal,* [mɛɾəl] is heard and for *medallion,* [mədæljən]. What underlying form must be posited once the child recognizes that the morpheme MEDAL occurs in both words?

Assuming the child has recognized MEDAL as a common element in *metal* and *metallic* and MEDAL as a common element in *medal* and *medallion,* here are the pronunciations that have been observed:

Metal		**Medal**	
[mɛɾəl]	[mət^{h}ælək]	[mɛɾəl]	[mədæljən]
metal	metallic	medal	medallion

The child could account for the different pronunciations of the morpheme metal by positing /mɛtæl/ in the lexicon and applying rules that change this underlying form into the realized surface forms. Focusing first on the consonants, an underlying form /mətæl/ would require a phonological process that changed /t/ into [ɾ] to yield [mɛɾəl]. (Below you'll see why we've used underlying /æ/.)

For the word *medal* /mɛdæl/, this same rule will be needed to change /d/ into [ɾ] in [mɛɾəl]. The tapping rule formulated earlier would change underlying /t/ and /d/ into [ɾ] between a stressed vowel and an unstressed vowel. Using a somewhat more formal notation, the rule would be:

Tapping Rule

alveolar stop	→	tap	/	V +stress	___ V −stress

Slips of the Tongue

In a tribute to England's Queen Victoria, the warden of one of Oxford's colleges aimed to toast "our dear old queen" but is reputed instead to have offered "Three cheers for our queer old dean!" Ascribed to the Reverend William Archibald Spooner, that slip of the tongue and many other speech errors continue to amuse and fascinate observers. But speech errors like those are of scientific value and can help illuminate the structure of the mental lexicon. Accidentally calling the "dear old queen" a "queer old dean" suggests that the brain/tongue combination has access not only to whole words but to separate sound segments. In "queer old dean," the onset [d] of the syllable [di:r] 'dear' has been exchanged with the onset [kw] of the syllable [kwi:r]. Both syllables have the same rhyme, probably helping to prompt the confusion and suggesting that syllable rhymes are also available to the brain. Other speech errors offer insight into which parts of words besides segments and onsets are accessible to the brain and therefore susceptible to "slips of the tongue."

A collection of speech errors collected and analyzed by linguist Victoria A. Fromkin point definitively to anticipated sound *segments*, as in an example like "alsho share" [ɔlʃo ʃɛr] for *also share* and "reek-long race" for *week-long race*, where an anticipated segment is substituted for a preceding one. Other examples also point to the psychological reality of whole sound segments, as in "teep a cape" for *keep a tape* and "the nipper is zarrow" for *the zipper is narrow*. Whole segments need not be consonants; they may be vowels, as in *Wang's bibliography* mistakenly called "Wing's babliography" and *dissertation topic* mistakenly pronounced [dɪsərtɑʃən tepək]. Strong evidence pointing to the availability of segments comes from examples like "frish gotto" for *fish grotto* and "blake fruid" for *brake fluid*. In "frish gotto," [r] has been extracted from the [gr] cluster and inserted after [f] to create a cluster, while in "blake fruid" the second segment in each cluster has been exchanged for the other.

Evidence for access to onsets can be surmised from "brittle island in litany" for *little island in Brittany* and "coat thrutting" for *throat cutting*. The onset of the first syllable in *Brittany* has been interchanged with the onset of the first syllable in *little* ([br] and [l]), and the onset of *throat* has been exchanged with the onset of *cutting* ([θr] and [k]). Evidence for the reality of elements smaller than a segment comes from examples like "glear plue sky" for *clear blue sky*, where only a single feature of the segments is interchanged, while other features remain unchanged: voicing is interchanged, while place and manner of articulation remain unaffected. Thus, [p] in "plue" (like [b] in *blue*) remains a bilabial stop but assumes the voicelessness of [k] in *clear*, while [k] in *clear* assumes the voiced feature of [b] in *blue*. In another example, "pig and vat" instead of *big and fat*, only the voicing values of the segments [b] and [f] are interchanged, while their places and manners of articulation remain unaltered; thus *big* becomes "pig" and *fat* becomes "vat." Note that in all the speech errors cited here the sequence constraints of English are honored, and segments that don't exist in the language are not produced in error. Speech errors don't create nonpermissible sounds or sound sequences.

It's not surprising that phonological rules that account for one set of facts may also account for other facts. After all, phonological rules apply to *all* morphemes and words that meet the structural description unless they have been specifically blocked from applying. For instance, nouns like *tooth* and *foot* that have irregular plural forms are specifically blocked from taking the regular plural morpheme. If *tooth* and *foot* weren't marked as irregular, adults would say *tooths* and *foots* just as children do before they learn to exempt these morphemes from the regular processes.

Thus the relationship between the phonological representation of morphemes in the lexicon and their actual realization in speech is mediated by a set of processes that can be represented in phonological rules of significant generality. Not only will *metal* and *medal* be affected by the tapping rule, but so will every word that meets the conditions specified in the rule, including single-morpheme words like *butter, city, meter,* and *lady*; two-morpheme words like *writer, rider, raider, rooter*; and so on.

Vowels Consider a youngster who knows the words *photograph* and *photographer* ([forəgræf] and [fətʰɑgrəfər]). At some point the youngster posits the entry PHOTOGRAPH in the lexicon to represent the core of these two words. A moment's thought will suggest that an underlying form /fotɑgræf/ would represent the baseline knowledge needed for the pair of alternate realizations. Given the underlying representation /fotɑgræf/ and the surface forms [forəgræf] and [fətʰɑgrəfər] and the recognition that [ə] occurs only in unstressed syllables, a rule that changes unstressed vowels into [ə] would produce the correct vowels.

On the other hand, if /ə/ appeared in the underlying form, no rule could produce the correct surface forms, and here's why. In order to produce the [ɑ] in [fətʰɑgrəfər] from an underlying form /fətəgrəfər/, a rule would need to produce [ɑ] from underlying /ə/. But for the word *photograph*, different rules would be needed to produce [o] from underlying /ə/ in the first syllable and [æ] from underlying /ə/ in the third syllable. This amounts to knowing which vowels exist in the surface realizations and encoding that knowledge in the underlying form along with /ə/, but that is exactly what we assume does not happen. Instead, if we postulate different vowels in the underlying forms, a single rule can derive [ə] from any underlying vowel when it occurs in an unstressed syllable. We can now derive the pronunciations for these words by formulating the rule as follows:

$$\begin{matrix} \text{V} \\ -\text{stress} \end{matrix} \rightarrow [\text{ə}]$$

(An unstressed vowel is realized as schwa.)

Of course, a rule that relies on information about stress would require prior assignment of stress, a matter that lies beyond the scope of this chapter.

Until now, we've represented underlying forms essentially as a sequence of phonological segments—for example, /kæt/ for *cat* and /kæt + z/ for *cats*. As some of our rules have suggested, though, representations in terms of segments like /k/ and /t/ are shorthand for a set of phonological *features*, akin to the sequence of bracketed columns below, whose feature sets represent the segments /k/, /æ/, and /t/. Entries in the mental lexicon must embody information that is equivalent to the information in a feature matrix. (Some treatments of phonology in other books employ a partly different and more abstract set of features from those we use here.) In addition, as we saw earlier, they include information about syllable structure, in this case that /æt/ (shorthand for the features) constitutes the rhyme, /t/ the coda, and /k/ the syllable onset. In producing and comprehending speech, the brain is able to access several levels of phonological representation—not only segments and the features they comprise but syllable structure and more.

$$\begin{bmatrix} \text{-voice} \\ \text{velar} \\ \text{stop} \end{bmatrix} \begin{bmatrix} \text{low} \\ \text{front} \\ \text{vowel} \end{bmatrix} \begin{bmatrix} \text{-voice} \\ \text{alvelar} \\ \text{stop} \end{bmatrix}$$

Phonological Processes

In the course of examining phonological rules in this chapter, we noted examples of assimilation and insertion. Those and other phonological processes are common in the world's languages. Let's take a closer look.

Assimilation

Assimilation is a phonological process by which adjacent (and sometimes near-adjacent) segments become more alike. The process often reflects a kind of co-articulation, where the pronunciation of a sound segment anticipates some feature of an upcoming sound. We saw this with the English possessive and plural markers on nouns and the third-person singular marker on verbs. We posited a voiced alveolar fricative /z/ as the underlying form for each of those morphemes. When affixed to a particular stem, the underlying /z/ may be realized on the surface as [z] or [s], depending on the voicing of the final sound segment in the stem to which the affix is attached. For example, /z/ is realized as voiceless [s] in *racks* [ræks] because /ræk/ (the stem) ends in /k/, a voiceless segment; underlying /ræg + z/ *rags* is realized as [rægz] because the stem ends in /g/, a voiced segment; /z/ in /rip + z/ *reaps* is realized as [rips] following the final voiceless segment of the stem; voiced [z] in [ridz] *reads* follows the final voiced segment of the stem. Assimilation also accounts for the distribution of the past-tense realizations of /d/ as voiceless [t] following voiceless /s/ as in *kissed* [kɪst] and as voiced [d] after voiced /z/ as in *quizzed* [kwɪzd]. Assimilation of the voicing feature of the affixed morpheme to the specific voicing feature of the final segment in the stem is rule-governed and entirely predictable in English.

Earlier we noted assimilation of Korean underlying /p/, which is realized as [b] between vowels; the underlying bilabial stop takes on the voicing feature of the surrounding vowels. German words whose underlying base ends in a voiced stop (as with the underlined parts of *Sta<u>b</u>* 'rod,' *Bun<u>d</u>* 'collar,' and *Ta<u>g</u>* 'day') are realized with a word-final *voiceless* stop. If, however, those words carry inflections so that the final consonant of the base is followed by a vowel, the underlying /b/, /d/, and /g/ are realized as the *voiced* stops [b], [d], and [g].

Besides assimilation of the voicing feature, sounds can assimilate to other sounds in place or manner of articulation. Consider the common pronunciation of English *utmost* /ʌtmost/ as [ʌpmost]. In anticipation of the upcoming *bilabial* /m/, the alveolar stop /t/ is realized as a *bilabial* stop [p], thereby assimilating to a bilabial place of articulation. In the pronunciations of *unkind* as [əŋkaɪnd] and *inclusive* as [ɪŋklusəv], the underlying alveolar nasal /n/ of UN- and IN- is realized as a *velar* [ŋ], anticipating the *velar* /k/ that follows; in pronouncing *unpopular* as [əmpɑpjələr] the underlying alveolar /n/ of UN- is realized as *bilabial* [m], anticipating the *bilabial* /p/ ahead.

Besides voicing assimilation and assimilation of place of articulation, one sound segment may assimilate to another in manner of articulation. Consider relaxed pronunciations of expressions like *good night* /gʊdnaɪt/ as [gʊnnaɪt] or [gʊnaɪt], where the alveolar stop /d/ is realized as the nasal [n], thus becoming

Try It Yourself What is the explanation for the fact that so many English words containing the prefix UN- (with underlying form /ən/ or /ʌn/)—words like *unaware*, *unopened*, *untidy*, *undamaged*, *unregistered*, *unhelpful*, *unlawful*, and *unjust*—maintain in their surface realization the alveolar place of articulation of the underlying nasal?

more similar to /n/ in manner of articulation. A more complex case can be exemplified in words like *congratulations* /kəngrætjuleʃənz/ pronounced as [kəngræʤəleʃənz], where the voiceless /tj/ is realized as the voiced affricate [ʤ]. Besides in voicing, assimilation has occurred in manner of articulation. Recall from the previous chapter that affricates are basically stop-fricatives. In *congratulations*, the sequence /tj/ is surrounded by vowels (which are voiced, of course). But note, too, that /tj/ involves a stop plus approximant. We described approximants in the previous chapter as characterized by a closure not sufficient to cause the characteristic friction of a consonant. In going from a stop through the approximant to the vowel, /t/ and /j/ assimilate to [ʧ], which is voiced to [ʤ] between the vowels /æ/ and /u/.

We noted in Chapter 3 that Spanish intervocalic /b/ is realized as [β]. The underlying /b/ is a stop, whereas vowels are continuous sounds. It is understandable, then, that an underlying stop /b/ might take on the continuous feature of the fricative [β], as in *cabo* 'end' pronounced [kaβo], and the same assimilation of manner can be heard intervocalically in *nudo* [nuðo] 'knot,' *lago* [laɣo] 'lake,' and similar Spanish words.

Try It Yourself The informal spellings *wanna*, *gonna*, and *hafta* capture assimilation in casual pronunciations of *want to*, *going to*, and *have to*. Focusing on whether the assimilation is in voicing or place of articulation or manner of articulation, describe what has been assimilated to what in each.

Deletion

Underlying consonants, particularly when they occur in clusters, are sometimes deleted. Consider the pronunciation of the English words *clothes* as [kloz] (not [kloðz]), *Arctic* as [ɑrtɪk] (not [ɑrktɪk]), *sandwich* as [sænwɪʧ] (not [sændwɪʧ]). In some cases, what appears to be deletion could also be viewed as assimilation, as with *granted* pronounced as though it were "granned" and *candidate* as though it were "canidate." Several Modern English words have spellings that reflect older pronunciations, from which a consonant has been deleted, as in *knife*. The /k/ has also been deleted from *knight*, along with the sound represented by <gh>. Earlier, the English word *ring* was spelled *hring* and pronounced with an initial /hr/ cluster. A phonological process known as syncope deletes vowels, as in the pronunciation of *federal* as [fɛdrəl] rather than [fɛdərəl], *interesting* as [ɪntrəstɪŋ] rather than [ɪntərɛstɪŋ], *mackerel* as [mækrəl] rather than [mækərəl].

Insertion

Sometimes a consonant cluster may be made easier to pronounce by inserting a vowel, thereby simplifying the cluster. English examples include pronouncing *triathlon* as [traɪæθəlɑn] rather than [traɪæθlɑn]; *disastrous* as [dɪzæstərəs] rather than [dɪzæstrəs]; and *ambidextrous* as [æmbidɛkstərəs] rather than [æmbidɛkstrəs].

Metathesis

The phonological process that reorders sounds within words, often making them easier to pronounce, is called metathesis. Common pronunciations like *cavalry* /kævəlri/ as [kælvəri] and *prescription* /priskrɪpʃən/ as [pərskrɪpʃən] illustrate the process in English.

From Lexical Entries to Surface Realizations: What the Brain Knows

Let's reexamine the phonological processes from this and an earlier chapter in order to determine which phonological units the brain must access for those processes to operate. For assimilation (for example, of the plural morpheme /z/ to a stem ending in a voiceless consonant as in *cats* [kæts] or *peaks* [piks]), the relevant phonological unit is the *feature* voice. Likewise for nasalization, aspiration, and lengthening, where the relevant unit again is the phonological *feature*—in these cases, the set of features that constitute a *natural class* (nasals, voiceless stops, voiced consonants). For reduplication (discussed in Chapter 2), the relevant unit is the *syllable*. For rhyming, as in poems, the relevant phonological unit is the *rhyme*, a structural part of the syllable. (For alliteration, as mentioned earlier in the chapter in the section "Syllables and Syllable Structure," the relevant unit is the *onset*, another structural part of the syllable.) In sum, ordinary phonological processes indicate that the brain can access segments, features, syllables, and structural parts of syllables.

Computers and Phonology

Several decades ago researchers thought it would be a matter of only a few years until computers would be able to recognize speech and synthesize it. (Think of *speech recognition* as turning speech into print and of *speech synthesis* as turning print into speech.) The process has taken longer than anticipated, and the reasons don't lie in limitations of technology. Despite the fact that children master the phonology of their language at a young age, linguists have not yet nailed down the extraordinary complexity of the phonological processes that characterize human languages. Phoneticians, phonologists, psycholinguists, and others have not yet sufficiently modeled what humans do when we produce and understand utterances. Natural speech occurs in a continuous stream and cannot be segmented without knowledge of the particular language involved. Just how children and adults segment a stream of spoken language into distinct words and recognize the sound segments in those words isn't fully understood.

The synthesis of speech by machine has also proved challenging. To understand why, consider the string of sounds that occur in a simple word like *sand*. It might appear straightforward to put together a machine-generated form of /sænd/: produce a voiceless alveolar fricative /s/, then the vowel /æ/, then an alveolar nasal /n/, and finally an alveolar stop /d/, and put them together with no silence in between. But notice, to begin with, that the vowel of *sand* differs markedly from the "same" vowel in *hat*: in *sand* it's nasalized; in *hat* it isn't. If a synthesizer produced the vowel of *hat* in the word *sand*, it would sound artificial. Therein lies one challenge for speech synthesis: how to blend sounds into one another as people do. In ordinary human speech, there is no separation between the sound segments of a word.

But the situation is more complicated than that. A sound is a bundle of features. The phonological form of *sand* isn't just the segments /s, æ, n, d/ but also the features characterizing each segment:

/s/	/æ/	/n/	/d/
-voice	+voice	+voice	+voice
alveolar	low front	alveolar	alveolar
fricative	vowel	nasal	stop

It's important to recall that the articulation of each feature in a segment does not start and end at the same time as all the other features in that segment. The voicelessness of /s/ doesn't abruptly end and the voicing of /æ/ start at exactly the same millisecond as the fricative character of /s/ stops and the vowel character of /æ/ begins. The systems of the vocal tract move continuously in the production of even the simplest words, and it is useful to conceptualize a word as a series of vocal gestures in continuous movement from one gesture to the next.

To make synthesized speech sound natural, a good deal more about the nature of underlying phonological forms, their surface realizations, and their pronunciations, as well as about how to get from the brain to vocal articulation (and from perceived speech to the brain), must be understood. ■

Summary

- Phonology is the study of the sound systems of languages.
- A phoneme is a unit in the sound system of a language. It is an abstract element, a set of phonological features (for example, bilabial, stop) having several language-specific predictable realizations (called allophones).
- Two words can differ minimally by virtue of having a single pair of different phonemes (as in *pin/bin* or *tap/tab*).
- Each phoneme comprises a set of allophones. Each allophone is the specific rule-governed and therefore predictable realization of the phoneme in a particular linguistic environment.
- The allophones of a phoneme never contrast but occur in complementary distribution or free variation. Allophones of the same phoneme cannot signal the sole difference in a minimal pair of words with different meanings.
- Two languages can have the same sounds but structure them differently within their systems. Both Korean and English have the sounds [p], [p^h], and [b]. In English, unaspirated [p] and aspirated [p^h] are allophones of the phoneme /p/, while [b] belongs to the phoneme /b/. In Korean, by contrast, [p^h] and [p] are

separate phonemes (they contrast), while [b] is the allophone of the phoneme /p/ that occurs between voiced sounds.

- Each simple word in a speaker's lexicon consists of a sequence of phonemes that constitutes its underlying phonological representation. Underlying forms differ from pronunciations and cannot generally be observed in speech directly.
- From the underlying form of a word, the phonological rules of a language specify the allophonic realizations of underlying segments in accordance with their linguistic environment.
- In acquiring a language, children must uncover the phonological rules of their language and infer efficient, economical underlying forms for all lexical entries. Given these underlying forms, the phonological rules of a language will specify the rule-governed features of the surface form.
- Phonological rules (in one model of phonology) may be ordered with respect to one another, with the first applicable rule applying to the underlying form to produce a derived form and subsequent rules applying in turn to successive derived forms until the last applicable rule produces a surface form. Two dialects of a language may contain some of the same rules but apply them in different orders, thereby producing different surface forms.
- Within a word, sounds are organized into syllables, and syllables have constituent parts, including a rhyme and an onset. Every syllable must have at least a rhyme; the rhyme must have at least a nucleus, and it may contain a coda.
- Languages have sequence constraints on the structure of permissible syllable types and the occurrence of particular consonants and vowels within syllable types.
- CV is the most common syllable type in the world's languages. English has an unusually large range of syllable types, including clusters of two and three consonants. The particular consonants that can appear in each position of a cluster are limited or constrained.
- Stress is contrastive in English: "Not every white house is the White House."
- Phonological processes (for example, aspiration and tapping in English) can depend on syllable structure and stress, as well as on a sequence of sound segments.

What Do You Think? REVISITED

- *Crestfallen Kristen.* The cabdriver likely realizes Kristen is an English speaker because he hears the results of her unconscious application of English phonological rules to French words. For example, by a general rule English aspirates initial /p/ sounds, but French doesn't. Ditto for other voiceless stops. Given the many common French words beginning with a voiceless stop (*par, pas, pour, tout, très, tu, comme, croire, que*), there's plenty of opportunity for Kristen to cue her nonnative status by aspiration in a pattern the driver recognizes as English-based.
- *Jules Biker.* One explanation is that the /p/ in Jewel Spiker's family name is neutralized in the environment following /s/. In other words, because English does

not exploit the /p/-/b/ contrast in that environment, the sounds are pronounced more alike there than elsewhere. While initial /p/ in English is aspirated, initial /b/ is not. In the environment following /s/, as in *Spiker*, /p/ is not aspirated, and the absence of aspiration contributes to easy confusion of /p/ and /b/ in that environment.

- *Techie Tammy*. On the Internet, you quickly discover that speech can be readily synthesized sound for sound. You remember that, when you were in high school, you could easily recognize a synthesized voice. Now, many times, you still can recognize synthesized speech, but sometimes it's more natural than you recall in the transition between sounds within a word and across words—precisely those environments where it previously sounded most artificial. You learn that much synthesized speech doesn't actually rely on putting individual sounds together but instead couples pairs of sounds, using the second half of the first one and the first half of the second one. It's as though *between* were made up of be + et + tw + we + en all squeezed together. You also learn that many of these combinations are already constructed and in reserve ready to be incorporated into a spoken response.
- *Spanish-speaking Simona*. English words allow for a variety of initial consonant clusters, including not only /sp/ and /st/ as in *speak* and *study* but even more complex ones like /str/ and /skw/ as in *stripe* and *squeeze*. Spanish has different sequence constraints and doesn't permit initial /sp/ or /st/ clusters. As a fan of Mexican cuisine you may be able to gain insight from names you know, and you quickly see that consonant clusters are uncommon. Some food names begin with a vowel (*arroz*, *empanada*, *enchilada*), but most begin with a single consonant: *pimiento*, *tapa*, *camarone*, *burrito*, *gazpacho*, *nacho*, *fajita*, *ceviche*, *chorizo*, *leche*, and *relleno*. Off the menu you know that Spanish words like *puerto* 'port' (as in *Puerto Rico*) and *puerco* 'pig' have initial clusters that English lacks. You conclude that Spanish may permit some initial clusters, but not /st/ or /sk/. Spanish speakers aiming to pronounce an initial consonant cluster with /s/ may face interference from the sequence constraints of their native tongue. One way to pronounce English words beginning with an /s-/ consonant cluster is to insert a vowel sound, creating a VC syllable like [ɛs], a syllable type permitted in Spanish.

Exercises

Practice Exercise

A. The words below are given in standard spelling and a broad phonetic transcription. Examine the allomorphic variation in the pronunciation of the underlined morpheme, and provide an underlying form from which the allomorphic variants could be derived by rule. Attend only to the *pronunciation* represented in the transcription, and ignore the spelling. (For this exercise, nasalization has been ignored.)

Example: metal [mɛɾəl] metallic [mət^{h}ælək] Underlying form: /mɛtæl/

human	[hjumən]	humanity [hjumænəɾi]
courage	[k^{h}ʌrəʤ]	courageous [k^{h}əreʤəs]
industry	[ɪndəstri]	industrial [ɪndʌstriəl]
medicine	[mɛɾəsən]	medicinal [mədɪsənəl]

B. Transcribe these monosyllabic words and underscore the rhyme in your transcription; then, in order, identify the nucleus, and (if they are present) the onset and coda; if any element is absent, write ∅.

Examples: rest: [rɛst] ɛ, r, st   clinched [klɪnʧt] ɪ, kl, nʧt

sit	spent	squirts
sin	squash	scrunched
scent	sprint	scratched

Based on English

4-1. Consider the following words with respect to how the sound represented by <t> is pronounced. For each column, specify the phonetic character of the allophone (how it is pronounced). Is it aspirated? Tapped? Then, as was done in this chapter for the allophones of English /p/, describe the allophones of /t/ and specify their distribution.

A	B	C	D
tougher	standing	later	petunia
talker	still	data	potato
teller	story	petal	return

4-2. Using the monosyllabic words below, provide a list of ten ordered pairs whose stress pattern indicates they are compounds—that is, with stress as in the examples. It will be helpful to mark the stress pattern on the vowel of each element, using ʹ for primary and ˋ for secondary stress.

Examples: tímezòne, shówhòrse

ball	beam	court	face	fall	free	gear	hand	hat
heart	hold	horse	house	kick	lance	land	lap	life
light	paint	port	rein	ride	road	show	style	table
throw	tide	time	top	way	weight	year	zone	

4-3. The following words appear not to exist in English. Some could exist, but others violate the sequence constraints of English. Identify the potential words, and explain why the others are not permitted. For the potential words, provide an appropriate spelling in the standard orthography.

pɛtribɑr	twɪntʃ	rizənənt
pʌpkəss	blɪbjulə	læktomæŋgjuleʃən
pæŋgəkt	spret	spwənt

4-4. **a.** Make a list of as many monosyllabic English words as you can, each of which represents a different onset of three consonants. *Example*: *spr* in *spread*

b. Examine the initial clusters you listed in (a) and answer these questions:

Which consonants can occur first in a three-consonant cluster onset?

Which consonants can occur second in a three-consonant cluster onset?

Which consonants can occur third in a three-consonant cluster onset?

Examine your three lists to decide whether or not they constitute natural classes, and provide the name for any that do constitute a natural class.

4-5. Although English relies on the contrast between /p/ and /b/ (*pill* vs. *bill*), it doesn't exploit the contrast in all environments. For example, following /s/ (as in *spell* and *spin*), no contrast is possible. Hence, there is no pair of words such as /sbɪn/ and /spɪn/. When a language exploits a distinction in some environments but not all, the potential contrast tends to be neutralized where it isn't exploited. As a consequence, the /p/ of *pill* differs more from the /b/ of *bill* than does the /p/ of *spin* (try distinguishing "spin" from "sbin"). For one thing, the /p/ of *spin* (but not the /p/ of *pill*) lacks aspiration, like the /b/ of *bill*. Thus a feature that distinguishes /p/ and /b/ elsewhere is not exploited following /s/.

Below are two sets of words. Those in column I contain a contrast that English exploits in that environment but not in the environment of column II. In other words, for the words in column II there cannot be a contrast based on the sound difference represented in the pair of words in the same line in column I.

	I	II
i.	sit seat	sing ring king
ii.	bit beat	here beer peer
iii.	hat hate	hang sang rang
iv.	tad dad	sting star study
v.	cad gad	skill score scam

a. Identify the segment that is likely to prompt different phonetic transcriptions and specify what those transcriptions would be.

b. Characterize the environment (in column II) that supports the neutralization.

c. Based on your knowledge of English phonology (such as its sequence constraints), provide reasons for preferring one of the transcriptions over the other.

4-6. We noted earlier that the English tapping rule could be reformulated in terms of syllables instead of vowel segments. Formulate the tapping rule in terms of syllables and their parts.

4-7. For the words below, identify each syllable's rhyme and nucleus and (where appropriate) onset and coda. *Example*: for *past*, rhyme: *ast*; nucleus: *a*; onset: *p*; coda: *st*

twin turned e-vil love-lorn a-tro-cious re-spec-tive

4-8. Examine the careful and casual pronunciations of the English expressions below; then for each one, (a) provide a transcription of the careful pronunciation; (b) identify the name of the phonological process that links them; (c) describe what actually occurs in the process in each particular case. Use one of these as identifiers for the process: Assimilation (ASS); Deletion (DEL); Insertion (INS); Metathesis (MET). An example is given for the first expression. (For this exercise, you may ignore vowel changes.) (*Note*: /ɱ/ is the IPA symbol for a labiodental nasal.)

Expression	Careful	Casual	Process	Details
athlete	æθlit	æθəlit	INS	schwa inserted at syllable boundary between /θ/ and /l/, perhaps for ease of articulation
emphasis	ɛmfəsɪs	ɛɱfəsəs		
nuclear	nukliər	nukjələr		
espresso	ɛsprɛso	ɛksprɛso		
memory	mɛməri	məmri		
prostate	prɑstet	prɑstret		
pass him	pæs hɪm	pæsəm		
won't you	wont ju	wontʃə		

Based on Languages Other Than English

4-9. Fijian has prenasalized stops among its inventory of phonemes. The prenasalized stop [^{n}d] consists of a nasal pronounced immediately before the stop, with which it forms a single sound unit. Consider the following Fijian words as pronounced in fast speech:

vi^{n}di	'to spring up'	dina	'true'
ke^{n}da	'we'	dalo	'taro plant'
tiko	'to stay'	vu^{n}di	'plantain banana'
tutu	'grandfather'	ma^{n}da	'first'
viti	'Fiji'	tina	'mother'
dovu	'sugarcane'	mata	'eye'
do^{n}do	'to stretch out one's hand'	mokiti	'round'
		veve^{n}du	(a type of plant)

On the basis of these data, determine whether [d], [^{n}d], and [t] are allophones of a single phoneme or constitute two or three separate phonemes. If you find that two of them (or all of them) are allophones of a single phoneme, give the rule that describes the distribution of each allophone. If you analyze all three as separate phonemes, justify your answer. (*Note*: In Fijian all syllables end in a vowel.)

4-10. Examine the following words of Tongan, a Polynesian language. (*Note*: In Tongan all syllables end in a vowel.)

tauhi	'to take care'	sino	'body'
sisi	'garland'	totonu	'correct'
motu	'island'	pasi	'to clap'
mosimosi	'to drizzle'	fata	'shelf'
motomoto	'unripe'	movete	'to come apart'
fesi	'to break'	misi	'to dream'

a. On the basis of these data, determine whether [s] and [t] are allophones of a single phoneme in Tongan or are separate phonemes. If you find that they are allophones of the same phoneme, state the rule that describes where each allophone occurs. If you conclude that they are different phonemes, justify your answer.

b. In each of the following Tongan words, one sound has been replaced by a blank. This sound is either [s] or [t]. Without more knowledge of Tongan than you could figure out from the preceding question, is it possible to make an educated guess as to which of these two sounds fits in the blank? If so, provide the sound; if not, explain why.

___ ili	'fishing net'	fe ___ e	'lump'
___ uku	'to place'	lama ___ i	'to ambush'

c. In the course of the last century, Tongan borrowed many words from English and adapted them to fit the phonological structure of its words.

kaasete	'gazette'	suu	'shoe'
tisi	'dish'	koniseti	'concert'
sosaieti	'society'	pata	'butter'
salati	'salad'	suka	'sugar'
maasol	'marshall'	sikaa	'cigar'
sekoni	'second'	taimani	'diamond'

How does the phonemic status of [s] and [t] differ in borrowed words and in native Tongan words? In other words, is the situation the same in these borrowed words? Write an integrated statement about the status of [s] and [t] in Tongan. (*Hint*: Your statement will have to include information about which area of the Tongan vocabulary each part of the rule applies to.)

4-11. The distribution of the sounds [s] and [z] in colloquial Spanish is represented by the following examples in phonetic transcription:

izla	'island'	ʧiste	'joke'
fuersa	'force'	eski	'ski'
peskado	'fish'	riezgo	'risk'
muskulo	'muscle'	fiskal	'fiscal'
sin	'without'	rezvalar	'to slip'
rasko	'I scratch'	dezde	'since'
resto	'remainder'	razgo	'feature'

mizmo	'same'	beizbɔl	'baseball'
espalda	'back'	mas	'more'

Are [s] and [z] distinct phonemes of Spanish or allophones of a single phoneme? If they are distinct phonemes, support your answer. If they are allophones of the same phoneme, specify their distribution.

4-12. Consider the following Russian words. On the basis of this limited list, where does Russian appear to have a contrast between [t] and [d] and where does it appear not to have one? (*Note*: An apostrophe marks a palatalized consonant.)

pərʌxot	'steamboat'	t'ɛlə	'body'
gəz'ɛtə	'newspaper'	pot	'perspiration'
zapət	'west'	dərʌgoj	'dear'
rat	'glad'	d'ɛlə	'business'
zdan'ijə	'building'	ʃtat	'state'
most	'bridge'	pot	'under'

4-13. In Samoan, words may have two forms, one called "bad speech" (used in informal oratory when addressing peers or kin) and another called "good speech" (used in literary and religious situations and with foreigners). The difference between the two forms can be described by phonological rules. (*Note*: The Samoan words for "good" and "bad" do not carry the same connotations in this case as the English words.)

"bad"	**"good"**	
taatou	kaakou	'us all'
teine	keiŋe	'girl'
taŋata	kaŋaka	'man'
ŋaŋana	ŋaŋaŋa	'language'
totoŋi	kokoŋi	'price'
nofo	ŋofo	'to stay'
ŋaalue	ŋaalue	'to work'
fono	foŋo	'meeting'

d. Describe the phonological difference between the "bad" and "good" forms. Which is more basic—the "good" form or the "bad" form? (In other words, which one can serve as the underlying form for both forms?)

e. Wherever possible, fill in the blanks in the following table. If it is impossible to know the form of a missing word, explain why.

"bad"	**"good"**	
manu	______	'bird'
mate	______	'dead'
______	maŋoo	'shark'
______	kili	'fishing net'
tonu	______	'correct'
______	kaŋi	'to cry'

4-14. In German, the sequence of letters <ch> can represent (among other things) either of two sounds: [ç] (a voiceless palatal fricative) or [x] (a voiceless velar fricative). On the basis of the following data, determine whether these two sounds are distinct phonemes or allophones of a single phoneme.

kɛlç	*Kelch*	'cup'
fıçtə	*Fichte*	'fir tree'
knœçl	*Knöchel*	'knuckle'
kɔx	*Koch*	'cook'
tsurɛçt	*zurecht*	'in good order'
vʊxt	*Wucht*	'weight'
çırʊrk	*Chirurg*	'surgeon'
nüçtʊrn	*nüchtern*	'sober'
bux	*Buch*	'book'
bəraıç	*Bereich*	'scope'
hɛkçən	*Häkchen*	'apostrophe'
bax	*Bach*	'brook'

If [ç] and [x] are distinct phonemes, justify your answer. If they are allophones of the same phoneme, specify their distribution.

4-15. In this chapter you learned that Japanese sequence constraints allow syllables of the forms CV, V, and (when the second C is a nasal) CVC. Using that information, divide the words given in the Japanese vowel chart (Table 3.5) into syllables: *ima* 'now'; *aki* 'autumn'; *buji* 'safe'; *yoru* 'to approach'; *sensei* 'teacher.' Now do the same for the borrowed words *beesubooru* 'baseball' and *sutoraiku* 'strike,' where <ee> and <oo> represent long vowels, not doubled vowels.

4-16. In light of our discussions in this chapter and your experience with some of the preceding exercises, discuss the following quote from Halle and Clements (1983).

> The perception of intelligible speech is . . . determined only in part by the physical signal that strikes our ears. Of equal significance . . . is the contribution made by the perceiver's knowledge of the language in which the utterance is framed. Acts of perception that heavily depend on active contributions from the perceiver's mind are often described as illusions, and the perception of intelligible speech seems . . . to qualify for this description. A central problem of phonetics and phonology is . . . to provide a scientific characterization of this illusion which is at the heart of all human existence.

Especially for Educators and Future Teachers

4-17. As an exercise for a class of middle-school international students studying English, you've asked them to draw up a list of English names for games, and they offer these: *skokey, skwinty, twint, stwink, plopo, splopt, sprats, skretsht, spretched, skwickt, spwint, stwirl, tprash, stpop, frash, quirt, splast, plsats*. You recognize that a few names are not legitimate because they have sequences of *sounds* (not letters) that English doesn't permit. Which are impossible, and what explanation can you give the students about why they are impossible?

4-18. One of your students returns from a summer visit to Berlin, Paris, and Madrid and reports that when she listened to the local radio in those cities she could not separate the stream of speech into separate words: it all seemed a blur. English is different, she says, because English words are separate from one another and easy to identify. What would you tell her about the difference she experienced between her ability to hear English words and her inability to sort out those of German, French, and Spanish?

4-19. Recall from Chapter 3—and your own experience—that French speakers tend to pronounce the English word *thin* as "sin" and *this* as "zis." From this observation, what can you say about (a) the inventory of French consonants as compared to English ones; and (b) whether or not French uses voicing as a contrastive feature? Finally, what would you predict about how a French student might tend to pronounce the English words *then* and *thick*?

4-20. Focusing on high front vowels, carefully compare the Spanish vowel chart in Table 3.3 with the English vowel chart (inside front cover or Figure 3.4). Relying on those charts and any relevant experience of yours, identify with IPA symbols which pair of distinctive vowels in English you would predict to be challenging for Spanish-speaking students learning English, and explain why. Then cite two minimal pairs of English words (words that are identical except for those vowels) that could prove challenging for those students to perceive and produce.

4-21. Listening to your Persian-speaking students talking among themselves, you notice words borrowed from English that are pronounced in Farsi in systematically different ways from their original English pronunciation. For example, the word *professor* has been borrowed as *perofesor* and *studio* as *estudiyo*. You also notice that some words seem related to English words but are pronounced differently. In particular, you've noticed that the Farsi word for brother is *baradar*. What do these pronunciations suggest about sequence constraints on some word-initial consonant clusters in Persian?

4-23. Using phonological terms from this and the previous chapter, identify two characteristic features of "foreign accent" for students represented in the schools of your community. Aim to account for the differences between the way native and nonnative speakers of English pronounce certain accented words. It may help to reflect on (a) inventory of sounds, (b) phonological rules for the distribution of allophones, (c) sequence constraints for sounds.

Other Resources

Internet

- **LISU Website: http://www.CengageBrain.com** For users of this textbook. Provides updated Internet links as well as supplemental material for students and instructors. Here you will find interactive learning tools.

- **Speech on the Web: http://www.acoustics.hut.fi/~slemmett/speech.html** If you're interested in hearing synthesized speech, several websites provide examples. The site above is a "jump station," providing links to speech synthesizers and other valuable information. For some of them, you can type in something you wish to hear synthesized. Then, on a computer with multimedia capabilities, you can experience state-of-the-art text-to-speech synthesis. Other useful links worth exploring are given below:
  - **http://www.ims.uni-stuttgart.de/~moehler/synthspeech/**
  - **http://www.research.ibm.com/tts/**
- **Acapela Text to Speech Demo: http://www.acapela-group.com/text-to-speech-interactive-demo.html** Another site that allows you to type in what you want and hear it synthesized in several different male and female voices in Arabic, Danish, Dutch, English, and many other languages.
- **SpeechLinks: http://www.speech.cs.cmu.edu/comp.speech/SpeechLinks.html** This speech technology hyperlinks page contains hundreds of links to projects around the world. Besides links to technical papers (most beyond the reach of beginning students), you'll find links to sites exploring speech recognition and speech synthesis.
- **Ladefoged's Concatenative Speech Synthesis: http://www.phonetics.ucla.edu/vowels/chapter8/chapter8.html** This site provides American, English, and Scottish synthesized text-to-speech renditions of "The North Wind and the Sun were disputing which was the stronger, when a traveler came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveler take off his cloak should be considered stronger than the other."

Suggestions for Further Reading

- **Carlos Gussenhoven & Haike Jacobs. 2005. *Understanding Phonology*, 2nd ed.** (London: Hodder Arnold). An excellent follow-up to this chapter; rich and largely accessible.
- **Bruce Hayes. 2008. *An Introduction to Phonology*** (Malden, MA: Blackwell). An excellent and well-illustrated follow-up to this chapter, relying more heavily on abstract features than we have used here.
- **April McMahon. 2002. *An Introduction to English Phonology*** (Edinburgh: Edinburgh University Press). A basic, accessible treatment, including the phonology of words and phrases.
- **David Odden. 2005. *Introducing Phonology*** (Cambridge, UK: Cambridge University Press). A basic treatment, with separate chapters on phonetic transcription, allophonic relations, underlying representations, abstractness and psychological reality, and one devoted solely to analyses.

Advanced Reading

Clark et al. (2007), a basic textbook, largely accessible to readers who have mastered some phonetics and the phonology of this chapter, has provided some English words used as examples in this chapter. The "problem book" by Halle and Clements (1983) covers a broad range of languages and has an excellent introductory chapter going beyond what we have covered; it has chapters on complementary distribution, natural classes, phonological rules, and systems of rules. Roca and Johnson (1999) is an excellent workbook, with scores of problems in a range of languages; it can be used independently or as an accompaniment to *A Course in Phonology* by the same authors. Kaye (1989) is a lively, provocative, and mostly accessible follow-up to this chapter. More specialized treatments are available in Bybee (2002) and Goldsmith (1996). The speech error data are taken from Fromkin (1971), which has many more examples and is accessible at least to eager students.

References

Bybee, Joan. 2002. *Phonology and Language Use* (Cambridge, UK: Cambridge University Press).

Clark, John, Colin Yallop & Janet Fletcher. 2007. *An Introduction to Phonetics and Phonology*, 3rd ed. (Malden, MA: Blackwell).

Fromkin, Victoria A. 1971. "The Non-Anomalous Nature of Anomalous Utterances." *Language* 47, 1: 27–52.

Goldsmith, John A., ed. 1996. *The Handbook of Phonological Theory* (Malden, MA: Blackwell).

Halle, Morris & G. N. Clements. 1983. *Problem Book in Phonology* (Cambridge, MA: MIT Press).

Kaye, Jonathan. 1989. *Phonology: A Cognitive View* (Hillsdale, NJ: Erlbaum).

Roca, Iggy & Wyn Johnson. 1999. *A Course in Phonology* (Malden, MA: Blackwell).

———. 1999. *A Workbook in Phonology* (Oxford: Blackwell).