



## 3 The Sounds of Language

I take it you already know  
Of tough and bough and cough and dough?  
Others may stumble but not you  
On hiccough, thorough, lough and through.  
Well done! And now you wish, perhaps,  
To learn of less familiar traps?  
Beware of heard, a dreadful word,  
That looks like beard and sounds like bird.  
And dead: it's said like bed, not bead—  
For goodness sake don't call it "deed"!  
Watch out for meat and great and threat  
(They rhyme with suite and straight and debt).

T.S.W. quoted in Mackay (1970)

In Chapter 1, we noted some of the basic features of the human vocal tract and the intricate muscle interlacing in and around the mouth that give humans the ability to produce a wide range of sounds with great speed. Yet, as they chatter away, humans do not simply produce a random selection of these sounds. Only certain sounds are selected on a regular basis as significant for communicative activity. In order to identify and describe those sounds, we have to slow down the chatter of everyday talk and focus on each individual sound segment within the stream of speech. This may seem straightforward, but it is not an easy task.

## Phonetics

Fortunately, there is an already established analytic framework for the study of speech segments that has been developed and refined for over a hundred years and is known as the **International Phonetic Alphabet**, or **IPA**. In this chapter, we will look at how some of the symbols of this alphabet can be used to represent the sounds of English words and what physical aspects of the human vocal tract are involved in the production of those sounds. The full IPA chart can be found at [internationalphoneticalphabet.org](http://internationalphoneticalphabet.org).

The general study of the characteristics of speech sounds is called **phonetics**. Our main interest will be in **articulatory phonetics**, which is the study of how speech sounds are made, or articulated. Other areas of study are **acoustic phonetics**, which deals with the physical properties of speech as sound waves in the air, and **auditory phonetics** (or perceptual phonetics), which deals with the perception, via the ear, of speech sounds.

## Consonants

We are not generally aware of how we produce speech sounds and it takes a certain amount of concentration on what we are doing with our mouths to become capable of describing the individual sounds produced. We will begin with the consonants. When we describe the articulation of a consonant, we focus on three features: the voiced/voiceless distinction, the place of articulation and the manner of articulation.

## Voiced and Voiceless Sounds

To make a consonant sound, we start with the air pushed out by the lungs up through the trachea (or windpipe) to the larynx. Inside the larynx are your **vocal folds** (or vocal cords), which take two basic positions.

- 1 When the vocal folds are spread apart, the air from the lungs passes between them with no obstruction, producing **voiceless sounds**.
- 2 When the vocal folds are drawn together, the air from the lungs repeatedly pushes them apart as it passes through, with a vibration effect, producing **voiced sounds**.

The distinction can be felt physically if you place a fingertip gently on the top of your Adam's apple (i.e. that part of your larynx you can feel in your neck below your chin), then produce sounds such as Z-Z-Z-Z or V-V-V-V. Because these are voiced sounds, you should be able to feel some vibration. Keeping your fingertip in the same position, now make the sounds S-S-S-S or F-F-F-F. Because these are voiceless sounds, there should be no vibration. Another trick is to put a finger in each ear, not too far, and produce the voiced sounds (e.g. Z-Z-Z-Z) to hear and feel some vibration, whereas no vibration will be heard or felt if you make voiceless sounds (e.g. S-S-S-S) in the same way.

## Place of Articulation

Once the air has passed through the larynx, it enters the vocal tract and comes up via the pharynx, an extended tube shape about five inches (13 centimeters) long. It is then pushed out through the mouth (the oral tract) and/or the nose (the nasal tract). As noted in Chapter 1, we typically produce speech as we are breathing out and generally find it quite difficult to do very much talking while breathing in. Most consonant sounds are produced by using the tongue and other parts of the mouth to constrict, in some way, the shape of the oral tract through which the air is passing. The terms used to describe many sounds are those that denote the place of articulation of the sound: that is, the location inside the mouth at which the constriction takes place.

What we need is a slice of head. If we crack a head right down the middle, we will be able to see those parts of the oral cavity that are crucially involved in speech production. In Figure 3.1, in addition to lips and teeth, a number of other physical features are identified. To describe the place of articulation of most consonant sounds, we can start at the front of the mouth and work back. We can also keep the voiced-voiceless distinction in mind and begin using the symbols of the IPA for specific sounds. These symbols will be enclosed within square brackets [ ].

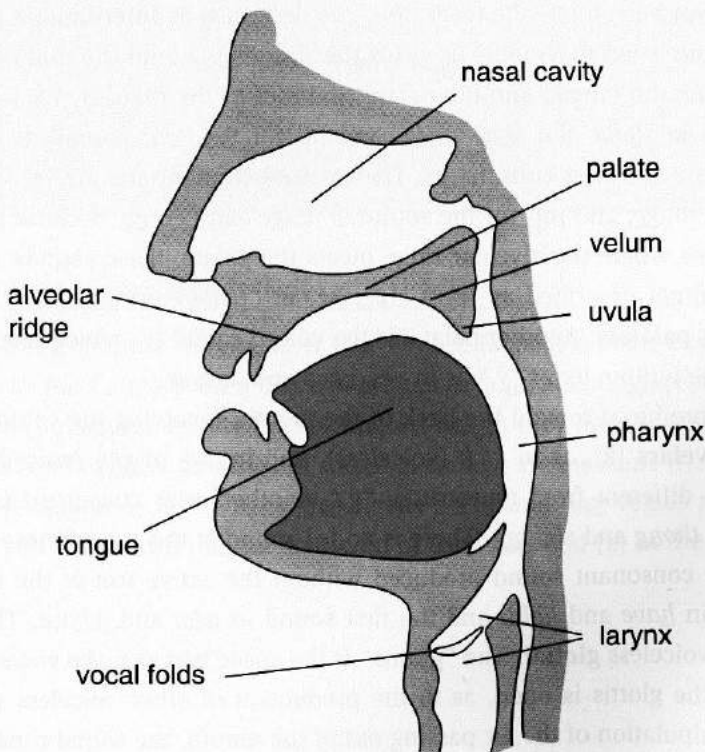


Figure 3.1

## Familiar Symbols

---

Many of the symbols used in phonetics to describe consonant sounds will be familiar. We use [p] for the voiceless consonant in *pop*. We use [b] in *Bob*, [m] in *mom* and [w] in *wet* for the voiced versions. These are **bilabial** consonants, made with both lips.

We use [f] and [v] for the **labiodentals**, which are formed using the upper front teeth and the lower lip at the beginning of *fat* and *vat*. The voiceless [f] is at the beginning and the voiced [v] is at the end of the pronunciation of *five*.

Behind the upper teeth is a rough area called the alveolar ridge. We raise the front of the tongue to this area when we make the **alveolar** sounds of [t] in *tot*, [d] in *dad*, [s], [z] in *size*, [r], [l] in *rail* and [n] in *nun*; [t] and [s] are voiceless, [d], [z], [r], [l] and [n] are voiced.

## Unfamiliar Symbols

---

Other symbols may be much less familiar, as in the two ways of representing the “th” sounds in English. We use [θ], called “theta,” for the voiceless version, as in *thin* and *wrath*, and at the beginning and end of the phrase *three teeth*. We use [ð], called “eth,” for the voiced version, as in *thus*, *then*, *feather* and *loathe*. Because the teeth are involved in creating these sounds, they are called **dentals**. If these sounds are made with the tongue tip between (= inter) the teeth, they are described as **interdentals**.

There are some special symbols used for the sounds made in the middle area of the mouth, involving the tongue and the palate (the roof of the mouth). We use [ʃ] for the “sh” sound, as in *shout* and *shoe-brush*, and [tʃ] for the “ch” sound, as in *child* and *church*. These are voiceless consonants. Their voiced counterparts are [ʒ] for the sound in *treasure* and *rouge*, and [dʒ] for the sound in *judge* and *George*. Because they are produced in an area where the alveolar ridge meets the palate, these sounds ([ʃ], [tʃ], [ʒ], [dʒ]) are sometimes described as “post-alveolar” or “palato-alveolar,” but we will just refer to them as **palatals**. Another palatal is the voiced sound [j], which often represents the sound of the written letter “y,” as in *yes*, *yoyo* and *lawyer*.

The sounds produced toward the back of the mouth, involving the velum, are represented by the **velars** [k], as in *kick* (voiceless), and [g], as in *gag* (voiced). Note that phonetic [g] is different from typewritten “g.” Another velar consonant is [ŋ], called “angma,” as in *thong* and *ringing*. There is no [g] sound at the end of these words.

There is one consonant sound produced without the active use of the tongue. It is the [h] sound in *have* and *hold*, and the first sound in *who* and *whose*. This sound is described as a voiceless **glottal**. The “glottis” is the space between the vocal folds in the larynx. When the glottis is open, as in the production of other voiceless sounds, and there is no manipulation of the air passing out of the mouth, the sound produced is [h].

A summary of the place of articulation for each consonant is presented in Table 3.1.



TABLE 3.1

Consonants	Voiceless	Voiced	Place of articulation
Bilabials	[p] <i><u>p</u>et, <u>t</u>ape</i>	[b], [m], [w] <i><u>b</u>et, <u>m</u>et, <u>w</u>et</i>	both (=bi) lips (=labia)
Labiodentals	[f] <i><u>f</u>at, <u>s</u>afe</i>	[v] <i><u>v</u>at, <u>s</u>ave</i>	upper teeth with lower lip
Dentals	[θ] <i><u>th</u>in, <u>b</u>ath</i>	[ð] <i><u>th</u>en, <u>b</u>athe</i>	tongue tip behind upper teeth
Alveolars	[t], [s] <i><u>t</u>op, <u>s</u>it</i>	[d], [z], [n], [l], [r] <i><u>d</u>og, <u>z</u>oo, <u>n</u>ut, <u>l</u>ap, <u>r</u>ap</i>	tongue tip to alveolar ridge
Palatals	[ʃ], [tʃ] <i><u>sh</u>ip, <u>ch</u>ip</i>	[ʒ], [dʒ], [j] <i><u>ca</u>sual, <u>g</u>em, <u>y</u>et</i>	tongue and palate
Velars	[k] <i><u>c</u>at, <u>b</u>ack</i>	[g], [ŋ] <i><u>g</u>un, <u>ba</u>ng</i>	back of tongue and velum
Glottals	[h] <i><u>h</u>at, <u>w</u>ho</i>		space between vocal folds

### Transcribing Sounds (Not Letters)

It is important to remember that written English is often a poor guide to pronunciation. We have already seen that words such as *bang* and *tongue* end with [ŋ] only, and there is no [g] sound despite the spelling. Although they have really different spellings, the first sound in *photo* and the last sound in *enough* are the same [f]. Perhaps more tricky are the final sounds in the pairs *face* versus *phase* and *race* versus *raise*: if you listen carefully, you will hear [s] in the first word of each pair and [z] in the second.

## Manner of Articulation

When we focus on the place of articulation for consonants, as in Table 3.1, we can see that [t] and [s] are similar in that they are both voiceless alveolars. But they are clearly different sounds. The difference is in how they are pronounced, or their manner of articulation. The [t] sound is a “stop” consonant and the [s] sound is a “fricative.” An analysis of the manner of articulation of English consonants is presented in Table 3.2. We should note that in some analyses, stops are described as “plosives” and what we call “glides” here may be described as “approximants” or “semi-vowels.”

TABLE 3.2

Consonants	Voiceless	Voiced	Manner of articulation
Stops	[p], [t], [k] <i>pet, talk</i>	[b], [d], [g] <i>bed, dog</i>	block airflow, let it go abruptly
Fricatives	[f], [θ], [s], [ʃ], [h] <i>faith, house, she,</i>	[v], [ð], [z], [ʒ] <i>vase, the, rouge</i>	almost block airflow, let it escape through a narrow gap
Affricates	[tʃ] <i>cheap, rich</i>	[dʒ] <i>jeep, rage</i>	combine a brief stop with a fricative
Nasals		[m], [n], [ŋ] <i>morning, name</i>	lower the velum, let air flow out through nose
Liquids		[l], [r] <i>load, light, road, write</i>	raise and curl tongue, let airflow escape round the sides
Glides		[w], [j] <i>we, want, yes, you</i>	move tongue to or from a vowel

## A Consonant Chart

Having described the most common consonant sounds used by English speakers, we can summarize the information in the following chart (Table 3.3). Along the top are the terms for place of articulation, as well as -V (voiceless) and +V (voiced). On the left-hand side are the terms for manner of articulation.

TABLE 3.3

	Bilabial		Labiodental		Dental		Alveolar		Palatal		Velar		Glottal	
	-V	+V	-V	+V	-V	+V	-V	+V	-V	+V	-V	+V	-V	+V
Stops	p	b					t	d			k	g		
Fricatives			f	v	θ	ð	s	z	ʃ	ʒ				h
Affricates									tʃ	dʒ				
Nasals		m						n				ŋ		
Liquids								l	r					
Glides		w										j		

### Glottal Stops and Flaps

Missing from Table 3.3 are two ways of pronouncing consonants that may also be heard in English, usually in casual speech situations. The **glottal stop**, represented by the symbol [ʔ], is produced when the space between the vocal folds (the glottis) is closed completely very briefly, then released. Many speakers produce a glottal stop in the middle of *Uh-uh* (meaning “no”), when they say the name *Harry Potter* as if it didn’t have the “H” or the “t,” or in the words *bottle* or *butter* without the “t” part.

If, however, you are someone who pronounces the word *butter* in a way that is close to “budder,” you are making a **flap**. It is represented by [ɾ]. This sound is produced by the tongue tip tapping the alveolar ridge briefly. Many American English speakers have a tendency to “flap” [t] and [d] consonants between vowels with the result that the pairs *latter/ladder*, *metal/medal* and *writer/rider* do not have distinct middle consonants. Those young students who were told about the importance of *Plato* in class and wrote it in their notes as *playdough* were clearly victims of a misinterpreted flap.

## Vowels

While the consonant sounds are mostly articulated via obstruction in the vocal tract, **vowel** sounds are produced with a relatively free flow of air. They are all typically voiced. To describe vowel sounds, we consider the way in which the tongue influences the shape through which the airflow must pass. To talk about a place of articulation, we think of the space inside the mouth as having a front versus a back and a high versus a low area. Thus, in the pronunciation of *heat* and *hit*, we talk about “high, front” vowels because the sound is made with the front part of the tongue in a raised position.

In contrast, the vowel sound in *hat* is produced with the tongue in a lower position and the sound in *hot* can be described as a “low, back” vowel. The next time you’re facing the bathroom mirror, try saying the words *heat*, *hit*, *hat*, *hot*. For the first two, your mouth will stay fairly closed, but for the last two, your tongue will move lower and cause your mouth to open wider. (The sounds of relaxation and pleasure typically contain lower vowels.)

We can use a vowel chart, like Table 3.4 (based on Ladefoged and Johnson, 2015), to help classify the most common vowel sounds in English.

TABLE 3.4

	Front	Central	Back
High	i		u
	ɪ		ʊ
Mid	e	ə	o
	ɛ	ʌ	ɔ
Low		æ	
		a	ɑ

### Front vowels

- [i] *bead, beef, key, me*
- [ɪ] *bid, myth, women*
- [e] *bed, dead, said*
- [æ] *bad, laugh, wrap*

### Central vowels

- [ə] *above, oven, support*
- [ʌ] *butt, blood, dove, tough*

### Back vowels

- [u] *boo, move, two, you*
- [ʊ] *book, could, put*
- [ɔ] *born, caught, fall, raw*
- [ɑ] *Bob, cot, swan*



## Diphthongs

In addition to single vowel sounds, we regularly create sounds that consist of a combination of two vowel sounds, known as **diphthongs**. When we produce diphthongs, our vocal organs move from one vocalic position [a] to another [ɪ] as we produce the sound [aɪ], as in *Hi* or *Bye*. The movement in this diphthong is from low toward high front. Alternatively, we can use movement from low toward high back, combining [a] and [ʊ] to produce the sound [aʊ], which is the diphthong repeated in the traditional speech training exercise [haʊ naʊ braʊn kaʊ]. In some descriptions, the movement is interpreted as involving a glide such as [j] or [w], so that the diphthongs we are representing as [aɪ] and [aʊ] may sometimes be seen as [aj] or [aw].

While the vowels [e], [a] and [o] are used as single sounds in other languages, and by speakers of different varieties of English, they are more often used as the first sounds of diphthongs in American English. Figure 3.2 provides a rough idea of how diphthongs are produced and is followed by a list of the sounds, with examples to illustrate some of the variation in the spelling of these sounds.

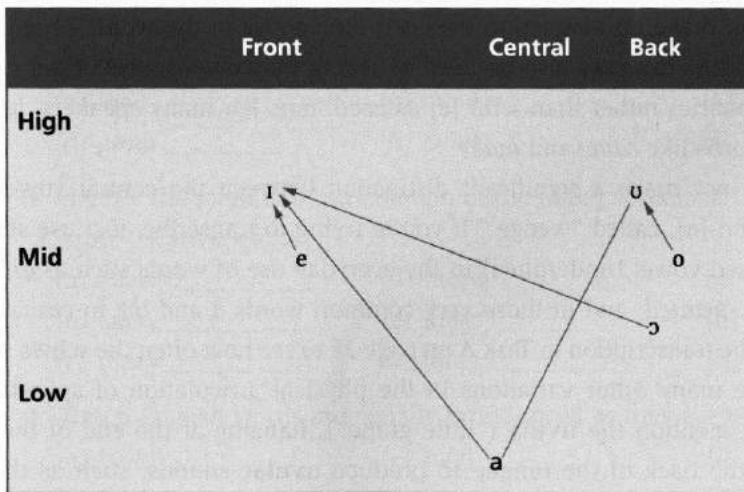


Figure 3.2

### Diphthongs

[aɪ] *buy, eye, I, my, pie, sigh*

[oʊ] *boat, home, owe, throw, toe*

[aʊ] *bough, doubt, cow*

[ɔɪ] *boy, noise, royal*

[eɪ] *bait, eight, great, late, say*

## American and British Diphthongs

The pronunciation of some diphthongs in Southern British English is quite different from North American English, as shown in Table 3.5. Note that the final [r] sound, typically pronounced in American varieties, is often omitted in Southern British English.

TABLE 3.5

	pair	peer	poor	pole
American	[peɪr]	[piːr]	[pʊr]	[pəʊl]
British	[peə(r)]	[piə(r)]	[pʊə(r)]	[pəʊl]

## Subtle Individual Variation

Vowel sounds are notorious for varying between one variety of English and the next, often being a key element in what we recognize as different accents. It may be, for example, that you make no distinction between the vowels in the words *caught* and *cot* and use [ɑ] in both. You may also be used to seeing the vowel sound of *pet* represented as [e] in dictionaries rather than with [ɛ] as used here. For many speakers, [e] is the short vowel in words like *came* and *make*.

You may not make a significant distinction between the central vowels [ə], called “schwa,” and [ʌ], called “wedge.” If you’re trying to transcribe, just use schwa [ə]. It is the unstressed vowel (underlined> in the everyday use of words such as *afford*, *collapse*, *photograph*, *wanted*, and in those very common words *a* and *the* in casual speech. You can check the transcription in Task A on page 38 to see how often the schwa sound occurs.

There are many other variations in the physical articulation of speech sounds. We didn’t even mention the **uvula** (“little grape”), hanging at the end of the velum. It is used with the back of the tongue to produce **uvular** sounds, such as the “r” sound, usually represented by [R], in the French pronunciation of *rouge* and *lettre*. The more we focus on the subtle differences in each sound, the more likely we are to find ourselves describing the pronunciation of a group or an individual speaker. Such differences help us to recognize a person’s voice as soon as he or she speaks. But those differences do not explain how we understand what total strangers with unfamiliar voices are saying. We are clearly able to disregard all the variation in the phonetic detail of voices and recognize each underlying sound type as part of a word with a particular meaning. To make sense of how we do that, we need to look at the more general sound patterns, or the phonology, of a language.