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).

Imagine the manager of a small restaurant, a man who has always had trouble with the spelling of unusual words, writing out a sign which he puts in the front window, advertising that they have a new seagh. You see the sign and you decide to ask what kind of new thing this is. When you hear the pronunciation, you recognize the word usually written as chef. How did he arrive at that other spelling? Well, it’s very simple, he says. Take the first sound of the word sure, the middle sound of the word dead, and the final sound of the word laugh. Isn’t that a seagh?
This tale, however unlikely, may serve as a reminder that the sounds of spoken English do not match up, a lot of the time, with letters of written English. If we cannot use the letters of the alphabet in a consistent way to represent the sounds we make, how do we go about describing the sounds of a language like English? One solution is to produce a separate alphabet with symbols that represent sounds. Such a set of symbols does exist and is called the phonetic alphabet. In this chapter, we will look at how these symbols are used to represent both the consonant and vowel sounds of English words and what physical aspects of the human vocal tract are involved in the production of those sounds.

**Phonetics**

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.

**Voiced and voiceless sounds**

In articulatory phonetics, we investigate how speech sounds are produced using the fairly complex oral equipment we have. 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 unimpeded. Sounds produced in this way are described as voiceless.
2. When the vocal folds are drawn together, the air from the lungs repeatedly pushes them apart as it passes through, creating a vibration effect. Sounds produced in this way are described as voiced.

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 comes up and out through the mouth and/or the nose. 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 cavity through which the air is passing. The terms used to describe many sounds are those which 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 you crack a head right down the middle, you will be able to see which parts of the oral cavity are crucially involved in speech production. 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 phonetic alphabet for specific sounds. These symbols will be enclosed within square brackets [ ].

![Diagram showing parts of the oral cavity and pharynx](image-url)
**Bilabials**

These are sounds formed using both (= bi) upper and lower lips (= labia). The initial sounds in the words *pat, bat* and *mat* are all **bilabials**. They are represented by the symbols [p], which is voiceless, and [b] and [m], which are voiced. We can also describe the [w] sound found at the beginning of *way, walk* and *world* as a bilabial.

**Labiodentals**

These are sounds formed with the upper teeth and the lower lip. The initial sounds of the words *fat* and *vat* and the final sounds in the words *safe* and *save* are **labiodentals**. They are represented by the symbols [f], which is voiceless, and [v], which is voiced. Notice that the final sound in the word *cough*, and the initial sound in *photo*, despite the spelling differences, are both pronounced as [f].

**Dentals**

These sounds are formed with the tongue tip behind the upper front teeth. The initial sound of *thin* and the final sound of *bath* are both voiceless **dentals**. The symbol used for this sound is [θ], usually referred to as “theta.” It is the symbol you would use for the first and last sounds in the phrase *three teeth*.

The voiced dental is represented by the symbol [ð], usually called “eth.” This sound is found in the pronunciation of the initial sound of common words like *the, there, then* and *thus*. It is also the middle consonant sound in *feather* and the final sound of *bathe*.

The term “interdentals” is sometimes used for these consonants when they are pronounced with the tongue tip between (= inter) the upper and lower teeth.

**Alveolars**

These are sounds formed with the front part of the tongue on the alveolar ridge, which is the rough, bony ridge immediately behind and above the upper teeth. The initial sounds in *top, dip, sit, zoo* and *nut* are all **alveolars**. The symbols for these sounds are easy to remember – [t], [d], [s], [z], [n]. Of these, [t] and [s] are voiceless whereas [d], [z] and [n] are voiced.

It may be clear that the final sounds of the words *bus* and *buzz* have to be [s] and [z] respectively, but what about the final sound of the word *raise*? The spelling is misleading because the final sound in this word is voiced and so must be represented
by [z]. Notice also that despite the different spelling of knot and not, both of these words are pronounced with [n] as the initial sound.

Other alveolars are the [l] sound found at the beginning of words such as lap and lit, and the [r] sound at the beginning of right and write.

**Palatals**

If you feel back behind the alveolar ridge, you should find a hard part in the roof of your mouth. This is called the hard palate or just the palate. Sounds produced with the tongue and the palate are called palatals (or alveo-palatals). Examples of palatals are the initial sounds in the words shout and child, which are both voiceless. The “sh” sound is represented as [ʃ] and the “ch” sound is represented as [ʧ]. So, the word shoe-brush begins and ends with the voiceless palatal sound [ʃ] and the word church begins and ends with the other voiceless palatal sound [ʧ].

One of the voiced palatals, represented by the symbol [ʒ], is not very common in English, but can be found as the middle consonant sound in words like treasure and pleasure, or the final sound in rouge. The other voiced palatal is [ʤ], which is the initial sound in words like joke and gem. The word judge and the name George both begin and end with the sound [ʤ] despite the obvious differences in spelling.

One other voiced palatal is the [j] sound used at the beginning of words like you and yet.

**Velars**

Even further back in the roof of the mouth, beyond the hard palate, you will find a soft area, which is called the soft palate, or the velum. Sounds produced with the back of the tongue against the velum are called velars. There is a voiceless velar sound, represented by the symbol [k], which occurs not only in kid and kill, but is also the initial sound in car and cold. Despite the variety in spelling, this [k] sound is both the initial and final sound in the words cook, kick and coke.

The voiced velar sound heard at the beginning of words like go, gun and give is represented by [ɡ]. This is also the final sound in words like bag, mug and, despite the spelling, plague.

The velum can be lowered to allow air to flow through the nasal cavity and thereby produce another voiced velar, represented by the symbol [ŋ], typically referred to as “angma.” In written English, this sound is normally spelled as the two letters “ng.” So, the [ŋ] sound is at the end of sing, sang and, despite the spelling, tongue. It occurs twice in the form ringing. Be careful not to be misled by the spelling of a word like bang – it ends with the [ŋ] sound only. There is no [ɡ] sound in this word.
Glottals

There is one sound that is produced without the active use of the tongue and other parts of the mouth. It is the sound [h] which occurs at the beginning of have and house and, for most speakers, as the first sound in who and whose. This sound is usually 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 that represented by [h].

Charting consonant sounds

Having described in some detail the place of articulation of English consonant sounds, we can summarize the basic information in the accompanying chart. Along the top of the chart are the different labels for places of articulation and, under each, the labels −V (= voiceless) and +V (= voiced). Also included in this chart, on the left-hand side, is a set of terms used to describe manner of articulation which we will discuss in the following section.

Limitations of the chart

This chart is far from complete. It contains the majority of consonant sounds used in the basic description of English pronunciation. There are, however, several differences between this basic set of symbols and the much more comprehensive chart produced by the International Phonetic Association (IPA). The most obvious difference is in the range of sounds covered.

We would go to an IPA chart for a description of the sounds of all languages. It includes, for example, symbols for the velar fricative sound you may have heard in the German pronunciation of the “ch” part of Bach or Achtung. It also includes sounds made with the
back of the tongue and the uvula (the “little grape” hanging at the end of the velum) which represents the “r” parts of the French pronunciation of *rouge* and *lettre*. Uvular sounds also occur in many native languages of North and South America. Other non-English sounds such as pharyngeals (produced in the pharynx) occur in languages such as Arabic. There are many other consonant sounds in the languages of the world.

Another way in which the chart is incomplete is the single entry covering “r” sounds in English. There can be a lot of variation among speakers in the pronunciation of the initial sound in *raw* and *red*, the medial sound in *very*, and the final sound in *hour* and *air*. Different symbols (e.g. [ɹ], [ʀ]) may be encountered in transcriptions where the different “r” sounds are distinguished.

Finally, in some phonetic descriptions, there are different symbols for a few of the sounds represented here. These alternatives are [s̥] for [ʃ], [z̥] for [ʒ], [c̥] for [ʧ], [j̥] for [ʤ] and [y] for [j]. For a fuller discussion of the use of these symbols, see Ladefoged (2006).

**Manner of articulation**

So far, we have concentrated on describing consonant sounds in terms of where they are articulated. We can also describe the same sounds in terms of how they are articulated. Such a description is necessary if we want to be able to differentiate between some sounds which, in the preceding discussion, we have placed in the same category. For example, we can say that [t] and [s] are both voiceless alveolar sounds. How do they differ? They differ in their manner of articulation, that is, in the way they are pronounced. The [t] sound is one of a set of sounds called stops and the [s] sound is one of a set called fricatives.

**Stops**

Of the sounds we have already mentioned, the set [p], [b], [t], [d], [k], [ɡ] are all produced by some form of “stopping” of the air stream (very briefly) then letting it go abruptly. This type of consonant sound, resulting from a blocking or stopping effect on the air stream, is called a stop (or a “plosive”). A full description of the [t] sound at the beginning of a word like *ten* is as a voiceless alveolar stop. In some discussions, only the manner of articulation is mentioned, as when it is said that the word *bed*, for example, begins and ends with voiced stops.

**Fricatives**

The manner of articulation used in producing the set of sounds [f], [v], [θ], [ð], [s], [z], [ʃ], [ʒ] involves almost blocking the air stream and having the air push through the
very narrow opening. As the air is pushed through, a type of friction is produced and the resulting sounds are called fricatives. If you put your open hand in front of your mouth when making these sounds, [f] and [s] in particular, you should be able to feel the stream of air being pushed out. The usual pronunciation of the word *fish* begins and ends with the voiceless fricatives [f] and [ʃ]. The word *those* begins and ends with the voiced fricatives [ð] and [z].

The sound [h], as in *Hi* or *Hello*, is voiceless and also usually included in the set of fricatives.

**Affricates**

If you combine a brief stopping of the air stream with an obstructed release which causes some friction, you will be able to produce the sounds [ʧ] and [ʤ]. These are called affricates and occur at the beginning of the words *cheap* and *jeep*. In the first of these, there is a voiceless affricate [ʧ], and in the second, a voiced affricate [ʤ].

**Nasals**

Most sounds are produced orally, with the velum raised, preventing airflow from entering the nasal cavity. However, when the velum is lowered and the air stream is allowed to flow out through the nose to produce [m], [n] and [ŋ], the sounds are described as nasals. These three sounds are all voiced. The words *morning*, *knitting* and *name* begin and end with nasals.

**Liquids**

The initial sounds in *led* and *red* are described as liquids. They are both voiced. The [l] sound is called a lateral liquid and is formed by letting the air stream flow around the sides of the tongue as the tip of the tongue makes contact with the middle of the alveolar ridge. The [r] sound at the beginning of *red* is formed with the tongue tip raised and curled back near the alveolar ridge.

**Glides**

The sounds [w] and [j] are described as glides. They are both voiced and occur at the beginning of *we*, *wet*, *you* and *yes*. These sounds are typically produced with the tongue in motion (or “gliding”) to or from the position of a vowel and are sometimes called semi-vowels.
In some approaches, the liquids [l], [r] and glides [w], [j] are combined in one category called “approximants.”

### Glottal stops and flaps

There are two common terms used to describe ways of pronouncing consonants which are not included in the chart presented earlier.

The **glottal stop**, represented by the symbol [ʔ], occurs when the space between the vocal folds (the glottis) is closed completely (very briefly), then released. Try saying the expression *Oh oh!*. Between the first *Oh* and the second *oh*, we typically produce a glottal stop. Some people do it in the middle of *Uh-uh* (meaning “no”), and others put one in place of “t” when they pronounce *Batman* quickly. You can also produce a glottal stop if you try to say the words *butter* or *bottle* without pronouncing the “-tt-” part in the middle. This sound is considered to be characteristic of Cockney (London) speech. (Try saying the name *Harry Potter* as if it didn’t have the “H” or the “tt.”) You will also hear glottal stops in the pronunciation of some Scottish speakers and also New Yorkers.

If, however, you are someone who pronounces the word *butter* in a way that is close to “budder,” then you are making a **flap**. It is represented by [D] or sometimes [ɾ]. This sound is produced by the tongue tip tapping the alveolar ridge briefly. Many American English speakers have a tendency to “flap” the [t] and [d] consonants between vowels so that, in casual speech, the pairs *latter* and *ladder* do not have distinct middle consonants. Nor do *writer* and *rider*, *metal* and *medal*. They all have flaps. The student who was told about the importance of *Plato* in class and wrote it in his notes as *play-dough* was clearly a victim of a misinterpreted flap.

This rather lengthy list of the phonetic features of English consonant sounds is not presented as a challenge to your ability to memorize a lot of terminology and symbols. It is presented as an illustration of how a thorough description of the physical aspects of speech production will allow us to characterize the sounds of spoken English, independently of the vagaries of spelling found in written English. There are, however, some sounds that we have not yet investigated. These are the types of sounds known as vowels and diphthongs.

### Vowels

While the consonant sounds are mostly articulated via closure or 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.)

The terminology for describing vowel sounds in English (e.g. “high front”) is usually based on their position in a chart, like the one shown here (based on Ladefoged, 2006), which provides a means of classifying the most common vowel sounds. Following the chart is a list of the major vowels with examples of familiar words illustrating some of the variation in spelling that is possible for each sound.

<table>
<thead>
<tr>
<th>Front</th>
<th>Front</th>
<th>Central</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>[i]</td>
<td>[ʊ]</td>
<td>[u]</td>
</tr>
<tr>
<td>Mid</td>
<td>[i]</td>
<td>[ə]</td>
<td>[o]</td>
</tr>
<tr>
<td>Low</td>
<td>[æ]</td>
<td>[æ]</td>
<td>[a]</td>
</tr>
</tbody>
</table>

![Figure 3.3](image)

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

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

**Back vowels**
- [ʊ] *boo, move, two, you*
- [o] *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 [ai], as in Hi or Bye. The movement in this diphthong is from low towards high front. Alternatively, we can use movement from low towards high back, combining [a] and [ʊ] to produce the sound [au], which is the diphthong repeated in the traditional speech training exercise [həʊ nəʊ braʊn kəʊ]. In some descriptions, the movement is interpreted as involving a glide such as [j] or [w], so that the diphthongs we are representing as [ai] and [au] may sometimes be seen as [aj] or [aw].

While the vowels [e], [a] and [o] are used as single sounds in other languages, and in some other varieties of English, they are only typically used as the first sounds of diphthongs in American English. The accompanying diagram 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.

<table>
<thead>
<tr>
<th>Front</th>
<th>Central</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Figure 3.4](image)

**Diphthongs**

- [ai] buy, eye, I, my, pie, sigh
- [au] bough, doubt, cow
- [ei] bait, eight, great, late, say
- [ou] boat, home, throw, toe
- [ɔɪ] boy, noise

**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. So, you may feel that some of the words offered in the earlier lists as examples don’t seem to be pronounced with the vowel sounds exactly as listed. Also, some of the sound distinctions shown here may not even be used regularly in your own speech. It may be, for example, that you make no distinction between the vowels in the words caught and cot and use [ə] 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.

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]. In fact, in casual speech, we all use schwa more than any other single sound. 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_.

There are many other variations in the actual physical articulation of the sounds we have considered here. The more we focus on the subtle differences in the actual articulation of each sound, the more likely we are to find ourselves describing the pronunciation of small groups or even individual speakers. Such subtle differences enable us to identify individual voices and recognize people we know as soon as they speak. But those differences don’t help us understand how we are able to work out what total strangers with unfamiliar voices are saying. We are clearly able to disregard all the subtle individual 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.