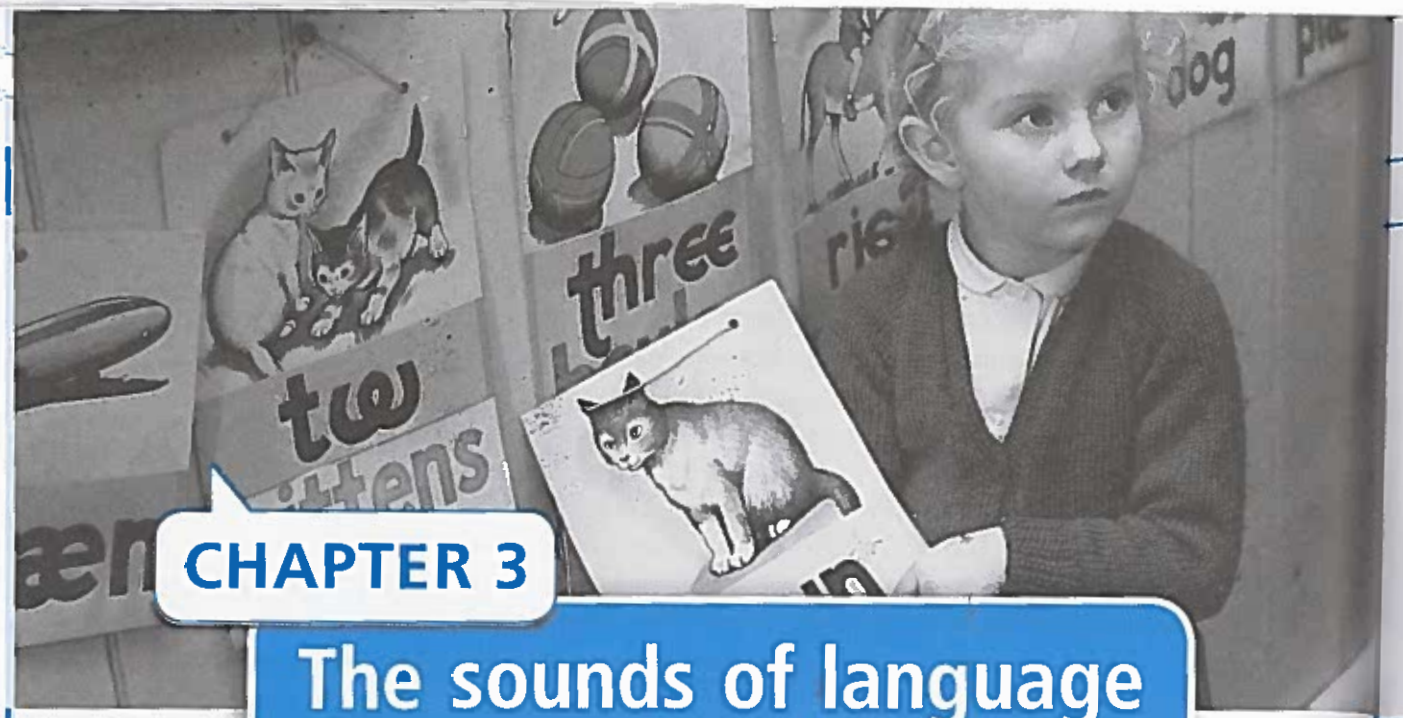

CHAPTER 3

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 the symbols of this alphabet can be 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.

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 that

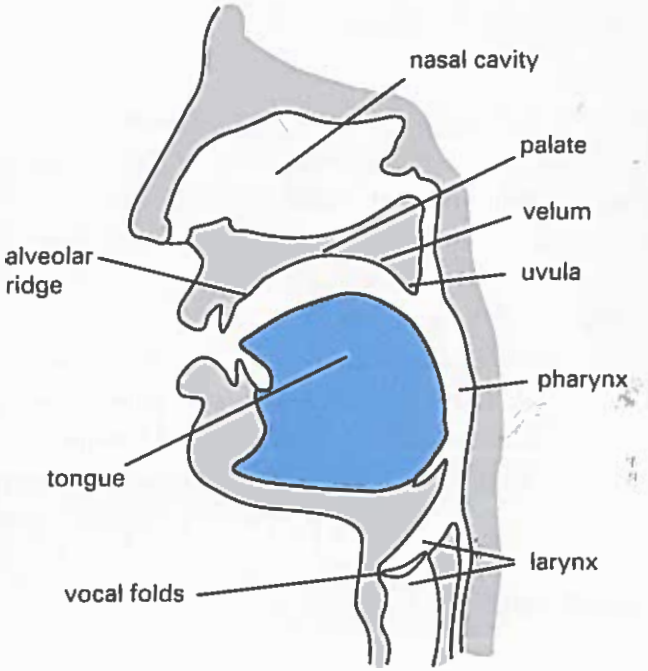

Figure 3.1

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 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 [].

Consonants

Familiar symbols

Many of the symbols used to describe consonant sounds will be familiar. We use [p] for the consonant in *pop*, [b] in *Bob*, and [m] in *mom*. These are **bilabial** consonants, made with both lips. We use [f] and [v] for the **labiodentals** (using upper teeth and lower lip) at the beginning and end of *five*. Behind the upper teeth is a rough area (the alveolar ridge) where we make the **alveolar** sounds of [t] in *tot*, [d] in *dad*, [s] and [z] in *size*, and [n] in *nun*.

Of course, there isn't always a match between written letters and phonetic symbols, as in the pronunciation of the sound at the beginning of *photo* and the end

of *enough*. In both cases, we would represent the sound with [f]. 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.

TABLE 3.1

Consonants	Place of articulation	Voiceless	Voiced
Bilabials	both (= bi) lips (= labia) together	[p] **p**at	[b], [m], [w] **b**at, **m**at, **w**et
Labiodentals	the upper teeth with the lower lip	[f] **f**at, sa**f**e	[v] **v**at, sa**v**e
Dentals	the tongue tip behind the upper teeth or between the teeth	[θ] **th**in, ba**th**	[ð] **th**en, ba**th**e
Alveolars	the front part of the tongue on the alveolar ridge (the rough area behind and above the upper teeth)	[t], [s] **t**op, si**t**	[d], [n], [z] [l], [r] **d**og, **n**ut, **z**oo l**a**p, **r**ap
Palatals	the tongue and the hard palate (on the roof of the mouth)	[ʃ], [ʧ] **sh**op, **ch**op	[ʒ], [ʤ], [j] casual, **g**em, **y**et
Velars	the back of the tongue on the velum (soft palate)	[k] **c**at	[g], [ŋ] **g**un, ba**ng**
Glottals	using the glottis, the open space between the vocal folds	[h] **h**at, **wh**o	

Unfamiliar symbols

Other symbols are 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 *three*, *wrath*. We use [ð], called "eth," for the voiced version, as in *thus*, *loathe*. Because the teeth are involved in the production of these sounds, they are called **dentals**, or in those cases where the tongue tip is between (= inter) the teeth, they may be 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*, *shoe-brush*, and [ʧ] for the "ch" sound, as in *child*, *church*. These are voiceless.

Their voiced counterparts are [ʒ] for the sound in *treasure*, *rouge*, and [ʤ] for the sound in *judge*, *George*. Another voiced sound made in this area is [j], which typically represents the "y" sound, as in *yes*, *yoyo*. Because the palate area is involved in these sounds, they are described as **palatals**.

The sounds produced toward the back of the mouth, involving the velum, are represented by the **velars** [k], as in *kick*, and [g], as in *gag*. Note that phonetic [g] is different from typewritten "g." We often use [k] to represent the sound of words beginning with "c," as well as some other letters, as in *cat*, *character* and *queue*.

One other consonant produced in this area is [ŋ], called "angma," as in *thong*, *ringing*. Be careful not to be misled by the spelling because both *bang* and *tongue* end with [ŋ] only. There is no [g] sound at the end of these words.

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

Consonants: manner of articulation

In Table 3.1, there is a detailed analysis of the place of articulation for consonants. From this we can see that [t] and [s] are similar in that they are both voiceless alveolars. But they're clearly different. 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."

Stops

In producing a stop consonant, we block the airflow briefly, then let it go abruptly. The voiceless stops are [p], [t], [k] and the voiced stops are [b], [d], [g]. So, the word *pet* begins and ends with voiceless stops and *bed* with voiced stops.

Fricatives

To produce a fricative, we almost block the airflow and force it through a narrow gap, creating a type of friction. The voiceless forms are [f], [θ], [s], [ʃ], [h], so that the word *fish* begins and ends with voiceless fricatives. The voiced versions are [v], [ð], [z], [ʒ], so the word *those* begins and ends with voiced fricatives.

Affricates

When we combine a brief stopping of the airflow with a release through a narrow gap, we produce the voiceless affricate [ʧ], at the beginning of *cheap*, and the voiced affricate [ʤ] at the beginning of *jeep*.

Nasals

Most sounds are produced orally, with the velum raised, preventing airflow from entering the nasal cavity. When the velum is lowered, allowing air to flow out through the nose, we can produce the nasals [m], [n] and [ŋ]. The words *morning*, *knitting* and *name* begin and end with nasals, all voiced.

Liquids

We describe the production of the two voiced sounds [l] and [r] as liquids. The [l] sound, as in *led* and *light*, is formed by letting the air flow around the sides of the tongue as the tip touches near the alveolar ridge. The [r] sound in *red* and *write* is formed with the tongue tip raised and curled back near the alveolar ridge.

Glides

The voiced sounds [w] and [j] are described as glides because they are produced with the tongue in motion (or "gliding") to or from the position of a vowel. The words *we*, *wet*, *yes* and *you* begin with glides (also called "semi-vowels").

A consonant chart

Having described the most common consonant sounds used by English speakers, we can summarize the information in Table 3.2. 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 3.2

	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									ʧ	ʤ				
Nasals		m						n				ŋ		
Liquids								l r						
Glides		w								j				

Glottal stops and flaps

Missing from Table 3.2 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 "tt," or even in the words *bottle* or *butter* without pronouncing the "tt" 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 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.)

We can use a vowel chart, like Table 3.3 (based on Ladefoged and Johnson, 2010), to help classify the most common vowel sounds in English.

TABLE 3.3

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

Front vowels	*Central vowels*	*Back vowels*
[i] *bead, beef, key, me*	[ə] *above, oven, support*	[u] *boo, move, two, you*
[ɪ] *bid, myth, women*	[ʌ] *butt, blood, dove, tough*	[ʊ] *book, could, put*
[ɛ] *bed, dead, said*		[ɔ] *born, caught, fall, raw*
[æ] *bad, laugh, wrap*		[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 towards 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.

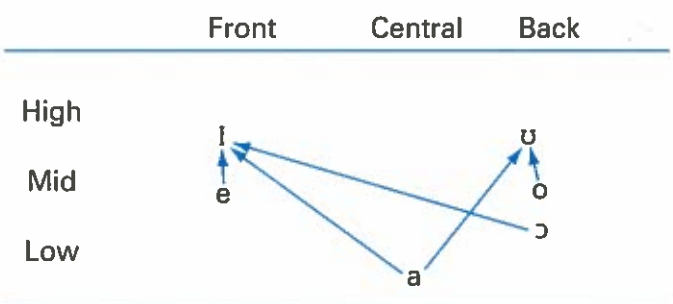

Figure 3.2

Diphthongs

[aɪ] *buy, eye, I, my, pie, sigh*	[oʊ] *boat, home, throw, toe*
[aʊ] *bough, doubt, cow*	[ɔɪ] *boy, noise*
[eɪ] *bait, eight, great, late, say*	

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 [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 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]. 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 *a̲fford*, *co̲llapse*, *photo̲graph*, *wanted̲*, and in those very common words *a* and *the̲*.

There are many other variations in the actual physical articulation of the sounds we have considered here. We didn't even mention the **uvula** (which means "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 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 will need to look at the more general sound patterns, or the phonology, of a language.