2

Sounds of Language: Phonetics and Phonology

This chapter deals with spoken language, and sets up the basic framework for describing speech sounds. The bulk of the chapter deals with the ways speech sounds are produced. It also explores the ways speech sounds pattern in the sound-systems of languages, which leads us to the notion of distinctive sounds or phonemes.

Chapter contents

Goals

The goals of the chapter are to:

- describe the basic structure of the speech organs or vocal tract;
- explain how speech sounds are produced;
- identify the main types of speech sounds and how they are classified;
- present the essentials of the main system for representing speech sounds, the International Phonetic Alphabet;
- outline the basic prosodic properties of speech (pitch, tone, intonation and stress) and how they are used in different languages;
- explain the notions of phoneme and allophone;
- show how to determine the phonemes of a language; and
- describe the major methods of transcribing speech.

Key terms

acoustic
allophone
articulatory phonetics
auditory phonetics
click
complementary distribution
consonant
diphthong
ejective
free variation
glottalic airstream
implosive
International Phonetic Alphabet (IPA)
intonation
manner of articulation
minimal pair
phone
phoneme
phonetics
pitch
place of articulation
pulmonic airstream
stress
suspicious pair
syllable
tone
velaric airstream
vocal tract
voicing
voice onset time
vowel

2.1 Fundamental properties of speech sounds

The speech chain

A simple and influential model of speech communication, the so-called speech chain model (presented in diagrammatic form in Denes and Pinson 1993: 5), identifies the following steps in conveying a message from speaker to hearer.

- A thought emerges in the brain of the speaker and is encoded in language.
- Messages are sent through nerves from the brain to the vocal apparatus – the muscles and organs that act together to produce speech sounds.
- The muscles and organs are positioned and set into motion.
- As a result, sounds are produced that travel through the air.
- These sounds reach the ears of the hearer, which 'process' the sounds, converting them into nervous signals.
- These signals travel along the auditory nerves to the brain of the hearer.
- The hearer's brain decodes these impulses, arriving at a thought (which is hopefully similar to the thought that started in the speaker's brain).
- The last three steps also apply to the speaker, in a feedback loop: the sound reaches the speaker's ear, is converted to electrical signals that travel to the brain, which decodes them and compares the spoken utterance with the intended utterance.

The three processes in the middle of this list, marked by the vertical bars, are the main concerns of phonetics, and give us the three primary divisions of the subject: **articulatory phonetics** (concerning the production of speech sounds), **acoustic phonetics** (concerning the physical properties of the sound waves) and **auditory phonetics** (concerning the perception of speech sounds). We deal mainly with articulatory phonetics in this chapter.

Phones

Readers of this book will be familiar with the idea that a stretch of speech such as a spoken version of *The farmer kissed the duckling* can be divided up into a sequence of **phones** or sounds coming one after another. First, there is the sound written *th* and pronounced in the same way as in ***th****ey* and ***th****em*; this is followed by another sound written *e* and pronounced like the *a* in *sof****a***. Then comes the *f* sound (as in ***f****rog*), an *a* sound as is also found in *f****a****ther*, and so on.

But the reality is not so simple. Figure 2.1 shows the sound wave for my production of this short sentence. What you can see immediately is that the sound is a continuous stream: it is not made up of separate blocks of sound separated by pauses. There is no precise point where you could say that the sound written *th* ends, and the one written *e* begins – the best we can say is that the *e* sound begins at around 0.06 seconds. Nor is there a clear division between the words, and the *f* sound runs on directly from the *e* sound at around 0.11 seconds, and is followed immediately by the *a* sound, at around 0.25 seconds. There is no interruption in the sound-stream throughout the word *farmer*. Later, between about 0.97 and 1.03 seconds there is a significant reduction in the sound wave. This might seem to be between *kissed* and *the*, but in fact it is within the word *kissed*! The final sound of this word, despite the spelling, is a *t*-sound, and that sound extends from about 0.91 to about 1.03, progressively becoming weaker. Similar remarks apply to the remainder of the utterance. In particular, the other apparent break in the sound wave – between 1.42 and 1.45 seconds – occurs within the *k* sound of *duckling*, not between the *k* sound and the following *l* sound.

You will have to take my word for the approximate divisions between the sounds. It will be impossible to verify them from Figure 2.1. In order to decide on the approximate transitions between the sounds I played the sound file many times while looking at the wave representation in a sound-analysis program called Praat (freely available at https://www.fon.hum.uva.nl/praat/). Even then there were uncertainties that I could only resolve by simultaneously examining another representation of the sounds, called a sound spectrogram. It is beyond the scope of the present chapter to discuss these acoustic representations. If you are interested in finding out more about spectrograms, you could consult Ladefoged and Disner (2012).

The same goes for the pronunciation of the sentence. The parts of the mouth involved in producing the sounds are in continuous motion; they do not move in a series of jerks from one fixed position to another. You do not first put your tongue in the position to make the *th* sound, then shift it instantaneously to the position to produce the *e* sound, then immediately put your lower lip against the upper teeth to make the *f* sound and so on.

Even if you cannot draw an exact boundary between each of the phones in the way that you can between the letters making up the words of the printed sentence (an ordinary handwritten version would raise similar difficulties of division), it does seem that the stream is made up of sequences of sounds of various types, following one after another.

The idea that you can divide stretches of speech into phones is a good approximation. To study the sounds of a language it is useful to represent phones in writing. For this purpose we use the **International Phonetic Alphabet** (IPA), a set of symbols based primarily on the Latin alphabet, and extensive and flexible enough to accommodate the sounds of any language. Table 2.1 shows the main symbols from the latest version of the IPA (dated 2015).

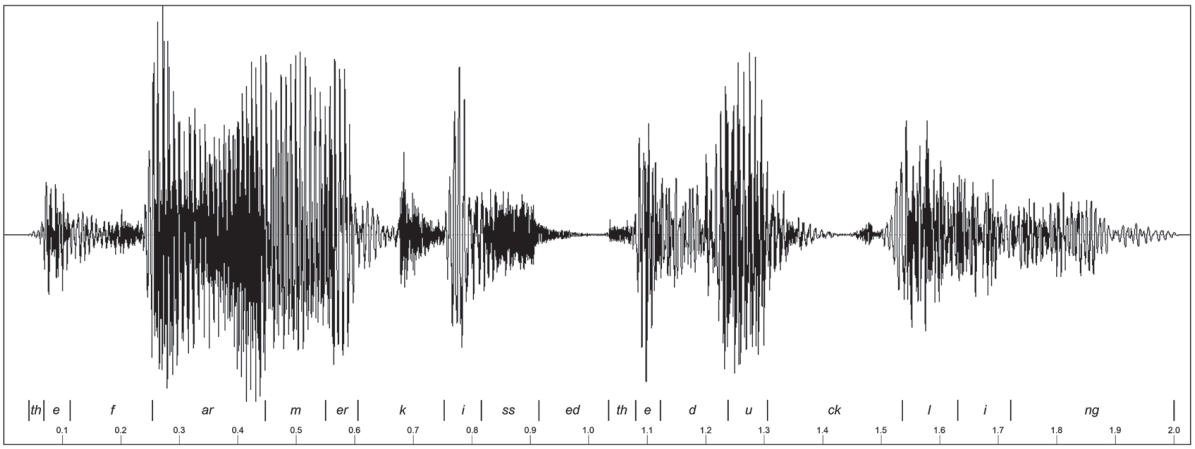

Figure 2.1 Sound wave of the author's production of *The farmer kissed the duckling*.

Table 2.1 Main symbols of the IPA (excerpted and slightly rearranged from IPA Chart, http://www.internationalphoneticassociation.org/content/ipa-chart, available under a Creative Commons Attribution-Sharealike 3.0 Unported License. Copyright © 2015 International Phonetic Association.)

Consonants (pulmonic)

		Bilabial	Labio-dental	Dental	Alveolar	Alveo-palatal	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Stop	voiceless	p			t		ʈ	c	k	q		ʔ
	voiced	b			d		ɖ	ɟ	g	ɢ		
Nasal		m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Fricative	voiceless	ɸ	f	θ	s	ʃ	ʂ	ç	x	χ	ħ	h
	voiced	ß	v	ð	z	ʒ	ʐ	ʝ	ɣ	ʁ	ʕ	ɦ
Affricate	voiceless				ʦ	ʧ						
	voiced				ʣ	ʤ						
Lateral	voiceless				ɬ*							
	voiced				l		ɭ	ʎ	ʟ			
Rhotic	tap, trill				ɾ, r		ɽ			ʀ		
	approximant				ɹ		ɻ					
Glides		w†						j	ɰ			

Notes
* ɬ is a voiceless lateral fricative; the other laterals (on the following line) are approximants.
† w is a voiced labio-velar glide.

Consonants (non-pulmonic)

Clicks	Voiced implosives	Ejectives
ʘ Bilabial	ɓ Bilabial	’ Examples:
ǀ Dental	ɗ̪ Dental-alveolar	p’ Bilabal
ǃ Retroflex	ʄ Palatal	t’ Dental-alveolar
ǂ Alveo-palatal	ɠ Velar	k’ Velar
ǁ Lateral	ʛ Uvular	s’ Alveolar fricative

Diacritics

n̥ d̥ Voiceless	b̤ Breathy voiced	tˠ Velarized
tʰ dʰ Aspirated	b̰ Creaky voiced	d̪ n̪ Dental
u̟ Advanced	tʷ Labialized	ẽ Nasalized
u̠ Retracted	n̩ Syllabic	dˀ Unreleased
ˈ Primary stress	ˌ Secondary stress	ː Long
ˑ Half long	. Syllable boundary	k͡p Coarticulated

Vowels			
	Front	Central	Back
High	i y	ɨ ʉ	ɯ u
	ɪ		ʊ
High Mid	e ø		ɤ o
		ə	
Low Mid	ɛ œ	ɜ	ʌ ɔ
	æ	ɐ	
Low	a		ɑɒ

Where symbols are paired, the first symbol indicates the unrounded vowel, the second the corresponding rounded vowel.

2.2 The vocal tract

The organs involved in producing the sounds of speech are referred to collectively as the **vocal tract**. These are the lungs, the larynx, the oral cavity and the nasal cavity; see Figure 2.2.

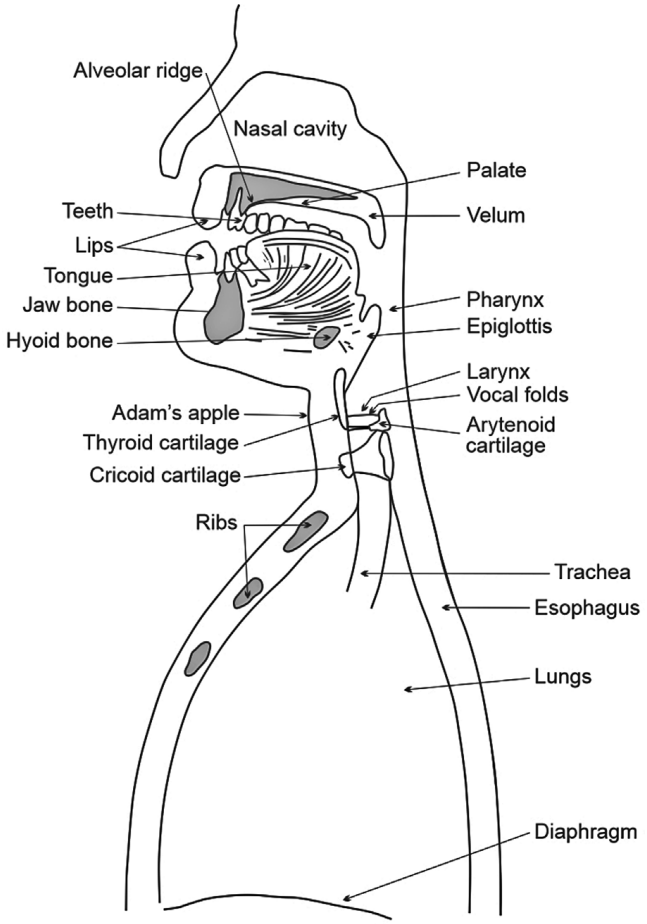

Figure 2.2 The human vocal tract.

The lungs

Most speech sounds are produced on a stream of air forced out from the lungs, through the trachea or windpipe, and then through the upper vocal tract, where the airstream is modified in various ways to produce different sounds. This stream of air is called an **egressive pulmonic airstream**. Speech in English and most other languages is usually produced on egressive pulmonic air.

It is also possible to produce speech sounds on air drawn into the lungs, on an **ingressive** pulmonic airstream. This is like speaking while breathing in. Although not as often used as egressive pulmonic air, in some languages it is used to convey emotional effects. For instance, in Danish and other Scandinavian languages, words – for example, *ja* 'yes' – are sometimes produced on an ingressive airstream to indicate sympathy or commiseration. Other airstream mechanisms used in human languages will be discussed in §2.4.

The larynx

In the larynx the airstream passes between a pair of muscular flaps called the **vocal folds** or **cords**, which can be drawn together to interrupt the airstream, or left open.

If you bring the vocal folds together closely but not too tightly, forcing air between them will cause them to vibrate regularly. This is how you produce *aaa*, the sound you might produce after the first sip of a long-awaited drink, or when lying down for a well-needed nap. These vibrations of the vocal folds, called **voicing**, can be felt by holding your thumb and first finger against your Adam's apple. Now say *ssss*, the sound of a deflating tyre. You will notice that the vibrations are not present while you produce this sound.

Sounds like *aaa* that are produced with voicing are called **voiced** phones, while those produced without it, like *ssss* are called **voiceless**. Compare *the* and *thin*. You will notice that the first phone of one of them is voiced, the other voiceless. Which is which?

In producing voiced speech sounds the vocal folds vibrate regularly, usually at between 80 and 400 cycles per second or **hertz** (Hz). This means that they move from the closed position to the open position and back again to the closed position between 80 and 400 times in a second, with each complete cycle taking about the same time.

While you are producing an *aaa* sound tighten your vocal folds further until the airstream is blocked off completely. Release it, continuing with the *aaa* sound. What you got in the middle is called a glottal stop, written ʔ in IPA. This sound can be heard in some English words – for example, the interjection written *uh-oh*.

From now on, we will use the IPA to represent phones, and enclose them in square brackets – thus [ʔ] is the glottal stop.

The oral cavity

After travelling through the larynx, the airstream passes through the **oral cavity** (mouth) and/or **nasal cavity** (nose), and then to the outside air. In the oral cavity the airstream can be modified in

many different ways to produce different sounds. The main organs used to make these modifications are the tongue and lips; the jaw is also involved, though in more subtle ways.

The phones [b] and [z] are produced on the same physical input, a voiced egressive airstream. But they sound quite different, and are formed differently. [b] is produced by first completely blocking the airstream through the oral cavity at the lips, allowing no air to escape to the outside; air from the lungs passes through the vibrating vocal folds and builds up behind the lip closure. Then the lips are parted and air behind the closure is released. [z] is produced by partially blocking the airstream from the lungs at the alveolar ridge and allowing the air to escape continuously with a noisy sound. These two phones differ in both the way they are produced and where they are produced.

The nasal cavity

At the back of the roof of the mouth is the **velum** or soft palate, which can be raised or lowered. When lowered, the airstream can enter the nasal cavity; when fully raised, the airstream is channelled into the oral cavity. In uttering the word *bad* [bæd] the velum remains closed throughout, and the air passes entirely through the oral cavity; all of these phones are **oral**. To produce the word *man* the only difference is that the velum is lowered so that the airstream enters into the nasal cavity; unless you speak very carefully, the velum will remain lowered throughout the word, including on the vowel, giving [mæ̃n]. (In the IPA, a ˜ over a symbol indicates that the sound is nasalized.) Phones like [m], [æ̃] and [n] are called **nasals**.

2.3 Types of phones

Speech sounds are divided into two main types, consonants and vowels. **Consonants** involve a constriction in the vocal tract, obstructing the flow of air; the airstream is impeded or interrupted somewhere along the path from the lungs to the outside. **Vowels** are produced with no significant obstruction to the passage of air through the oral cavity, and the air exits unimpeded through the oral cavity (and perhaps the nasal cavity as well). Vowels are the most resonant phones, those that resound or re-echo the most, like the chime of Big Ben. Consonants are either non-resonant (like [d] and [f]), or are somewhat resonant (like [m] and [l]), though these are less resonant than a vowel (compare these phones with [i] or [a]).

Consonants

Consonants are described in terms of the point where the airstream is impeded, and how it is impeded. These two properties are called the **place of articulation** and the **manner of articulation**. By convention, the places of articulation are shown across the top row of the IPA chart, the manners of articulation down the first column.

Places of articulation

The main places of articulation are illustrated schematically in Figure 2.3. Below we make a few remarks on each of them.

Labial

Labial sounds are made with the lips. If both lips are used, the phone is called a **bilabial**. The initial sounds of *bad* [bæd] and *mad* [mæd] are bilabials: they are made by bringing the upper and lower lips together; so also is the initial sound of *pad* [pʰæd].

Instead of bringing both lips together, you can bring the lower lip into contact with the upper teeth to form a **labiodental**. The phones [f] and [v], as in the beginning of the English words *fix* [fɪks] and *Vicks* [vɪks], are produced in this way.

Dental

Dental sounds are formed with the tongue and the upper teeth. Usually the tip of the tongue is used, and touches the upper teeth, as in English [θ] and [ð], the initial phones of *three* and *this*, respectively. In some dialects of English, including Californian English, the tip of the tongue

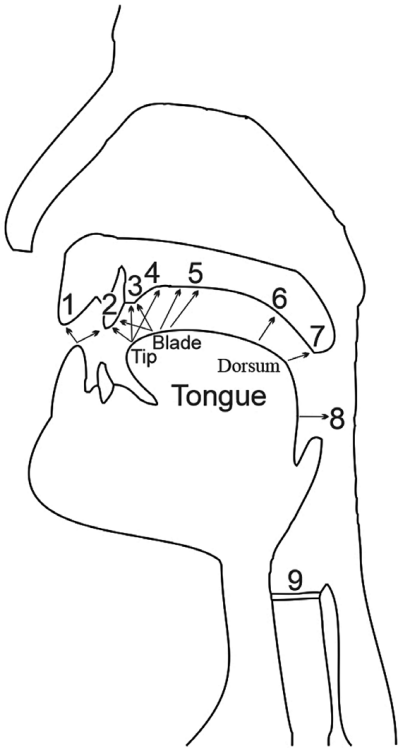

Figure 2.3 The upper vocal tract showing main places of articulation of consonants: 1, labial; 2, dental; 3, alveolar; 4, post-alveolar; 5, palatal; 6, velar; 7, uvula; 8, pharyngeal; 9, glottal. Arrows emanate from the main active parts (articulators) and point towards typical targets (passive articulators). More precise descriptions of places of articulation can be given by first specifying the active articulator, then the passive one, thus: apico-alveolar; lamino-palatal, dorso-velar and so on. (The tongue tip gives apico-; the blade gives lamino-; and the dorsum gives dorso-.)

protrudes between the upper and lower teeth for these consonants, which are called **interdentals**. The [n] and [d] phones in French are made with the tip of the tongue touching the upper teeth; this can be specified in the IPA by the tooth symbol under the letter, as in [d̪]. (In English the contact is slightly further back, at the alveolar ridge.)

In some languages dental sounds are made not with the tip of the tongue against the teeth, but the blade of the tongue, the part behind the tip. Many Australian languages have such phones. The Yindjibarndi (Pama-Nyungan, Australia) word *thugu* 'young boy', begins with this type of phone, as does *nhuurga* 'ankle'. (In Yindjibarndi spelling, the *h* indicates that it is the blade of the tongue that makes contact with the upper teeth.)

Alveolar

Sounds made with the front part of the tongue – usually the tip, sometimes the blade – touching or almost touching the alveolar ridge (the ridge on the roof of the mouth just behind the upper teeth) are called **alveolars**. The initial phones of the English words *top*, *dog*, *log*, *nag*, *rag*, and *sag* are all alveolars.

If the tip of the tongue makes contact just behind the alveolar ridge, the sound is sometimes called **post-alveolar** or **retroflex**.[1] Post-alveolar phones occur in the Yindjibarndi words *marda* 'blood' (the *rd* indicates that the *d*-sound is made further back than ordinary [d]) and *thurla* 'eye'.

Palatal

The palate or hard palate is the large region of the roof of the mouth extending from a little behind the alveolar ridge to the soft palate or velum. Sounds made with contact (or approximation) between the tongue and the palate are called **palatals**. If this is in the front part of the palate, immediately behind the alveolar ridge, the sound is an **alveo-palatal**. The English words *shingle*, *jungle* and *child* begin with alveo-palatals. For the first phones of the Hungarian words *nyak* 'neck' and *gyufa* 'match' (represented in spelling by *ny* and *gy*) the contact is a little further back, and these phones are palatals.

Velar

Behind the hard palate lies a soft area called the soft palate or velum. Consonants made with the back of the tongue touching the velum are called **velar** sounds. (Although it is physically possible to touch the velum with the tip of the tongue, it is not easy, and no language is known to employ this combination.) Velar sounds begin the English words *cull*, *kill*, *car*, *go*, *grip* and *give*. (The first three words begin with the voiceless velar stop despite the different spellings.) The nasal sound at the end of the English word *sing* is also a velar (IPA [ŋ]).

Uvular

The uvula is the appendage hanging down at the back of the velum. The relatively few phones made with the back of the tongue and the uvula are **uvulars**. The *r*-sound of Parisian French, many dialects of Dutch (Indo-European, Europe) and German, and a few dialects of English (such as Northumbrian English) are (not necessarily always) realized as uvular trills (IPA [ʀ]), made with rapid vibration of the uvula (see p. 37).

Pharyngeal

The pharynx is the chamber behind the back of the tongue, above the larynx, and roughly at right angles to the oral cavity. **Pharyngeal** consonants are made by pulling the root of the tongue back to narrow the pharynx so that the air passes through noisily. Pharyngeals are not found in English; Arabic (Afroasiatic, Arabian peninsula and North Africa), however, has them. So does Danish, in the *r*-sound of words like *råd* 'council'.

Glottal

Glottal consonants involve a constriction of the glottis or opening between the vocal folds. We have already encountered one segment made at this place, namely the glottal stop (see §2.2), which is produced by holding the vocal folds together and then releasing them. The initial phones of the English words *hot* [hɒt] and *hill* [hɪɬ] are also glottal.

Manners of articulation

Seven main manners of articulation are employed in human languages: stops, nasals, fricatives, affricates, laterals, rhotics and glides. In addition to these, the term *approximant* is used for phones that are produced by bringing the articulators towards one another, but not close enough to produce either a complete blockage or a turbulent noise.

Stops

A phone that has complete closure or blockage of the airstream is a **stop** or **plosive**. Stops can be made at (almost) any place of articulation, from the glottis ([ʔ]) to the lips (e.g. [p], [b]).

At the beginning of a word such as *bed* or *pet*, the bilabial phone involves complete blockage of the airstream, followed by an abrupt release of the pent-up air behind the blockage. When an English speaker says *bed*, the vocal folds are vibrating during the production of the initial stop. But for *pet* the vocal folds do not begin to vibrate until a short time after the first stop has been released. Stops like [b] are voiced; those like [p] are voiceless.

The period of time (measured in milliseconds or thousandths of a second) between the release of a stop and the onset of vibration of the vocal folds is called **voice onset time** (VOT). For the English [b] this is usually a negative number, as voicing begins before the release of the stop. For the English [p], at the beginning of a word, voice onset time is usually a little under 50 milliseconds (ms). For Danish, the voice onset times of stops are generally slightly longer, while in Spanish they are shorter.

When the VOT of a stop is rather long, as in the case of English and Danish voiceless stops, a puff of air follows the release of the stop, before the regular voicing of a following vowel begins. This is called **aspiration**, and indicated in IPA by a small raised ^{h}, as in [p^{h}]. You can observe this puff of air if you hold a small piece of paper loosely in front of your mouth when you say *pin*. Now say *spin*. What do you notice?

The stops at the end of the English words *bed* and *pet* are voiced and voiceless, respectively. Voicing extends into the closure of the alveolar stop in *bed*, but not (or more accurately, only a short way) into the closure of the alveolar stop in *pet*. The stop at the end of both words can be either **released** or **unreleased**: that is, you can either release or not release the pent-up air behind the

alveolar blockage, resulting in a noise. If you say *my pet dog Rover*, it is likely that you will not release the [t] of *pet*; but if you say this phrase very carefully, separating *pet* from *dog*, you might release the [t] before forming the following [d]. Unreleased stops are indicated in the IPA by a raised angle ˺. So the English word *pet* could be pronounced as either [pʰɛt] or [pʰɛt˺].

Nasals

Nasals have already been described (§2.2) as phones produced by lowering the velum to permit air to flow through the nose. Nasal consonants are like stops in having a complete blockage of the airstream through the oral cavity; but they allow the air to flow freely through the nasal cavity. Thus for [m] there is a complete closure of the oral cavity at the lips, as in [b] and [p]; but the velum is lowered, and the air travels out through the nose.

Nasals are normally voiced. Burmese (Sino-Tibetan, Myanmar), however, has voiceless bilabial, dental, palatal and velar nasals.

Fricatives

Fricatives are produced with incomplete closure at the place of articulation. A narrow passage is left open through which the airstream is forced, giving rise to a 'noisy' sound a bit like the sound you get by rubbing your hands on your clothing, or scratching your head.

Fricatives are found in many, though not all languages, and can be produced at any place of articulation, from the lips to the glottis. English has both voiced and voiceless fricatives. There is a voiced labiodental fricative, [v] as in *vale*, and a voiceless fricative at the same point of articulation, [f] as in *fail*. Other fricatives in English are dental [θ] and [ð], alveolar [s] and [z], alveopalatal [ʃ] and [ʒ], and glottal [h]. Ewe (Niger-Congo, Ghana) has both voiceless and voiced bilabial fricatives [ɸ] and [β], as well as labiodental fricatives [f] and [v]. Velar fricatives are found in German (as in *Buch* [buːx] 'book'), Old English, and many other languages including some dialects of English, including Scottish English (as in the word *loch* [ɬɔx] 'lake').

Affricates

Phones produced by combining a stop and a fricative are called **affricates**. English has two affricates, both alveopalatal: voiceless [ʧ] (the first and last sounds of *church*) and voiced [ʤ] (the first phone in *jungle*). The alveopalatal region is the most common place of articulation for affricates, though other places are possible. For instance, Mandarin Chinese has both aspirated and unaspirated alveolar and retroflex affricates, and Beembe (Niger-Congo, Democratic Republic of Congo) has labiodental affricates, again both aspirated and unaspirated.

Laterals

Phones like the initial [l] of English *laugh* are called **laterals** because the sides of the tongue are lowered at the point of articulation, allowing air to pass on both sides of a central closure. Sometimes laterals are produced with just one side of the tongue lowered.

The most common place of articulation of laterals is at the alveolar ridge. Dental and post-alveolar or retroflex laterals are also possible. A fair number of languages have palatal laterals, produced with contact between the blade of the tongue and the hard palate, with the sides open.

This sound is found in Castilian Spanish (Indo-European, Spain) ***ll****ave* 'key'. Velar laterals are also possible, though rare. In my dialect of English (Australian) words such as *milk* involve a velar lateral; the tip of the tongue is not used in this sound. The Papuan language Melpa (Papuan, New Guinea) also has a velar lateral.

In many dialects of English the back of the tongue is raised in the production of the lateral, especially when it occurs at the end of a word, as in *ball* or *school*; this lateral has a 'dark' sound quality. The IPA symbol for this sound is [ɬ]. In Australian English the alveolar lateral is usually dark, even at the beginning of words. By contrast, the lateral of Danish and French is always 'clear' in sound, and is produced without raising of the velum.

Like nasals, laterals are normally voiced; just a few languages have voiceless laterals. Burmese and Welsh (Indo-European, Wales) both have the voiceless alveolar lateral [ɬ]. It should be noted that this lateral is a lateral fricative, in contrast with the other laterals, which are not fricatives, but approximants (see note * to Table 2.1).

Rhotics

The term **rhotic** is used for *r*-like sounds, those phones represented in IPA by some variant of the Latin letter *r*. Rhotics come in a wide variety of shapes and forms, the main ones being taps, trills and approximants.

Taps (or flaps) involve a single rapid short closure (shorter than for a stop), usually between the tip of the tongue and the teeth or alveolar ridge, as in the Spanish words *caro* [karɔ] 'expensive' and *pero* [perɔ] 'but'. This is the usual rhotic of Scottish English.

Trills are phones consisting of two or more rapid taps one after another. Usually trills are produced with the tip of the tongue, as in Spanish *perro* [perɔ] 'dog'. They can also be produced by vibrating the uvula rapidly, as in Parisian French, Standard German and some varieties of Southern Swedish (Indo-European, Sweden) and Dutch. Bilabial trills – like the disrespectful 'raspberry' in English – are found in a few languages, for instance in Kele (Papuan, New Guinea).

Rhotic approximants involve proximity, rather than contact, at the place of articulation. The *r*-sound of most dialects of English is an approximant. However, it differs considerably between dialects. In Australian English and Estuary English (London) it is an apical approximant, in which the tip of the tongue points towards the alveolar ridge, but does not make contact with it. Uvular approximants are found in Northumbrian English (the uvula trill is sometimes also used in this dialect), and sometimes in German and French.

Glides

Glides or semivowels are the most vowel-like of the consonants, having the least constriction at the point of articulation. They are characterized by movement of one articulator, which travels towards but does not reach the other.

The *y*-sound of English, IPA [j], is a glide in which the blade of the tongue moves towards the palate. The blade of the tongue can also move towards the teeth. Bunuba (Bunuban, Australia) and Unggumi (Worrorran, Australia) have such glides (as well as the palatal glide). The so-called 'soft-*d*' of Danish (as in *mad* 'food') is a dental approximant; however, it involves movement of the tip rather than the blade of the tongue towards the teeth.

English has a second glide, the labiovelar [w], produced by moving the back of the tongue towards the velum and at the same time rounding the lips. This phone occurs in about three-quarters of the world's languages (the palatal glide is found in a slightly higher fraction of the languages, about 85 per cent). A few languages have a plain velar glide [ɰ], unaccompanied by lip rounding, and even fewer have a bilabial or labiodental glide without accompanying velarization.

Vowels

Vowels are speech sounds produced without interruption to the passage of air through the vocal tract. The vocal tract is used as a resonating chamber for an airstream vibrating from the action of the vocal folds; and as this suggests, vowels are normally voiced in all languages. The cavities above the glottis act rather like an organ pipe or the chamber of a wind instrument, except that it has a characteristic right-angled bend between the vertical pharynx and the horizontal oral cavity. The shape of the cavities can be modified by positioning and shaping the tongue in different ways, which has an effect comparable to directing the airstream in an organ into different sized pipes.

The position of the high point of the tongue during the production of a vowel effectively defines the size and shape of the two resonating chambers, and thus the quality of the vowel. If the high point is high up and towards the front of the mouth, you get vowel sounds like the one usually written *ee* in English, as in *beet* [biːt], or the one written *i* in *bit* [bɪt]. These vowels are called **high front** vowels. If instead the high point of the tongue is high up and towards the back of the oral cavity, you get a vowel like the one written *u* in *put* [pʰʊt]. These are **high back** vowels. For high vowels the body of the tongue is raised above its neutral or rest position.

If the body of the tongue rests at a relatively neutral height, as for the vowel of *bed* [bɛd] we have a **mid** vowel; lowering the body of the tongue results in a **low** vowel, such as the vowel of *fat* [fæt].

The front-to-back dimension is also usually divided into three: **front**, with the high point relatively front; **back**, with the high point towards the back; and **central**, with the high point in between, in the central region.

Table 2.2 shows the vowels of BBC English, the variety used by national newscasters in Britain, with the IPA symbols and illustrative examples.

Other dialects of English have slightly different ranges of vowels. General American English, for instance, the variety used by many national broadcasters in the USA, lacks the low back unrounded vowel [ɒ], and has [ɚ] in place of [ɜː]. There are also minor differences in the qualities of the vowels, due partly to differences in the positioning of the high point of the tongue.

New Zealand English has the high central vowel [ɨ] instead of the [ɪ] of other dialects.[2] This vowel, though not common, is found in various other languages, including Amharic (Afro-Asiatic, Ethiopia) and Nimboran (Papuan, West Papua).

High- and mid-back vowels are usually accompanied by lip-rounding, and so are called **rounded vowels**. High- and mid-front vowels are usually produced with spreading of the lips, while low vowels are usually produced with the lips in a neutral position.

Table 2.2 Chart of vowels of BBC English showing IPA representation and examples (based on Ladefoged and Disner 2012: 30).

	Front	Central	Back
High	i: *bead* ɪ *bit*		u: *boot* ʊ *put*
Mid	ɛ *bed*	ə *the* (as normally spoken) ɜ: *bird*	ɔ: *ought* ʌ *cut*
Low	æ *bat*		ɒ *cot*; ɑ: *bard*

These correlations are imperfect. Some languages have high- and/or mid-front rounded vowels. Danish has three rounded front vowels: high-front rounded [y]; mid-high front rounded [ø]; and mid-low front rounded [œ]. These have the same tongue positions as the three non-low front vowels [i], [e] and [ɛ],[3] respectively. Fewer languages have high- and/or mid-back unrounded vowels. But Vietnamese (Mon-Khmer, Vietnam) has three unrounded back vowels, [ɯ], [ɤ] and [ʌ], which have the same heights as the rounded [u], [o] and [ɔ].

The velum can be lowered during the production of a vowel, allowing air to pass through the nasal cavity. This gives a **nasal vowel**. We have already seen that vowels in English can be nasal before a nasal consonant, as in *man* [mæ̃n]. French has nasal vowels, as in *lent* [lã] 'slow'.

2.4 Some additional features

The previous section presented an overview of the basic types of speech sounds found in the world's languages. In this section we mention a few additional important features of the phonetics of human languages.

Airstream mechanisms

In addition to the pulmonic airstream, two other airstreams are used for the articulation of speech sounds: glottalic and velaric.

Glottalic

A number of languages of Africa, India and the Americas employ a **glottalic airstream** in the production of some (never all) phones. This airstream is created by closing the vocal folds and raising or lowering the larynx, while a closure – complete or partial – is made somewhere in the oral cavity.

Ejectives are produced on an egressive (outgoing) glottalic airstream, formed by raising the larynx so as to compress the air behind the oral closure; this closure is released while the glottal

closure remains, resulting in a popping sound. About a fifth of the world's languages have ejectives. Quechua (Quechuan, Peru) has three, with alveopalatal, velar and uvular places of articulation: [ʧʼaka] 'hoarse', [kʼujui] 'to twist' and [qʼaʎu] 'tomato sauce'.

Implosives are produced by pulling the larynx downwards during oral closure, and releasing the oral closure, resulting in an audible inrush of air. Only about 10 per cent of the world's languages have implosives. Sindhi (Indo-European, India) has bilabial (as in [ɓəni] 'field'), retroflex (as in [ɖĩnu] 'festival'), palatal (as in [ʃətu] 'illiterate') and velar (as in [ɠənu] 'handle') implosives, in addition to ejectives.

Velaric

The 'tutting' sound written *tsk tsk* (sometimes *tut tut*) in English involves an **ingressive velaric airstream**. It is produced by forming closures between the tip of the tongue and the alveolar ridge and the back of the tongue and the velum or uvula, keeping these closures in place while the body of the tongue is drawn downwards, rarefying the enclosed air. Then the contact between the tip of the tongue and the alveolar ridge is released, with the result that air is drawn into the mouth. The kissing sound is made in the same manner, except that front closure is at the lips. These sounds are called **clicks**.

So-called Khoisan languages and some neighbouring Bantu languages of south-east Africa employ the velaric airstream in speech sounds. !Xóõ (Taa) (Tuu, Botswana and Namibia) has bilabial, dental, alveolar, lateral and palatal clicks, along with a number of accompaniments including voicing, nasality, aspiration, ejection and so on. The five clicks, without accompaniments, are illustrated in: [ʘân] 'sleep', [ǀān] a type of shrub, [!āo] 'remain', [ǁâe] 'three' and [ǂám] 'return, go back (home)'.

Coarticulation

It was mentioned above that the *w*-sound of English is produced with simultaneous movement of the lips and of the back of the tongue towards the velum. Such sounds, with simultaneous constriction at two places of articulation, involve **coarticulation** or **double articulation**.

Stops and nasals can also be coarticulated. The most common are bilabial-velar combinations. Various languages of West Africa have such phones. Idoma (Niger-Congo, Nigeria) has coarticulated voiceless and voiced stops, and a coarticulated nasal: [àk͡pà] 'bridge', [àg͡bà] 'jaw' and [aŋ͡màa] 'body painting'. Other combinations are rarer. Yélî Dnye (Papuan, Rossel Island) also has bilabial-alveolar coarticulated stops and nasals, in addition to labio-velars, as in [t͡pɛnɛ] 'lung' and [n͡mo] 'bird'. Perhaps the most unusual coarticulated phone is found in Wari' (Chapacura-Wanham, Brazil): a voiceless dental stop coarticulated with a voiceless bilabial trill, [t͡ʙ̥].

Diphthongs

Vowels are produced with the tongue in a relatively steady state throughout their articulation. A **diphthong** is produced when instead the tongue is in constant motion throughout, travelling from

Table 2.3 Diphthongs of three dialects of English (based on Ladefoged and Disner 2012: 28, 30 and my own dialect)

BBC English	Australian English	General American English	Examples
[aʊ]	[æʊ]	[aʊ]	*bowed, loud*
[əʊ]	[oʊ]	[oʊ]	*grow, owe*
[ɔɪ]	[ɔɪ]	[ɔɪ]	*boy, quoit*
[eɪ]	[eɪ]	[eɪ]	*day, say*
[aɪ]	[aɪ]	[aɪ]	*my, dry*
[ɪə]	[ɪə]		*beer, leer*
[eə]	[ɛə]		*hair, lair*
[ʊə]	[ʊə]		*poor, boor*
[aə]			*hire, pyre*

one vowel position to another. English dialects generally have a fair number of diphthongs, as can be seen in Table 2.3. (There are additional minor differences in the beginning and end points of the diphthongs that are not shown.)

Syllables

Phones combine into larger units called **syllables**. Syllables are surprisingly difficult to characterize precisely, and there is disagreement among phoneticians concerning their definition. Nevertheless, syllables seem to have more psychological reality for speakers than do phones, and native speakers of any language are usually able to divide spoken utterances into syllables without difficulty.[4] For instance, the word *phonetics* has three syllables (the boundaries of which are indicated by dots), [fə.nɛ.tɪks]. Even if you are not a native speaker of a language it is not normally too difficult to guess what the syllable division of a word is. The Gooniyandi word *girili* 'tree' has three syllables, and you can probably guess what they are.[5]

Syllables generally consist of a vowel together with one or more consonants, usually before the vowel, sometimes after it. Some languages allow syllables consisting of just a vowel; some allow syllables consisting of a consonant, generally a nasal or lateral. English has syllables of both of these types: consider *about* [ə.bæʊt] and *medal* [mɛd.ɬ]; these words also illustrate syllables with a consonant before and after the vowel (or diphthong) – that is, CVC syllables (where C stands for consonant, and V for vowel or diphthong). English has a considerable number of syllable types (including CV, CVC, V, VC, CCV and C), and thousands of possible syllables.

Languages differ considerably in terms of the syllable types they allow. In Gooniyandi all syllables must contain a vowel; usually it is preceded by a consonant, giving the most common

form of the syllable CV; syllables can, however, also end in consonants. CV syllables are the most common syllables in human languages, and are the most frequent in most languages.

2.5 Prosodies

Vowels and consonants are segments that come one after another in speech (although the boundaries between them are generally not precisely delimited, as we have seen). Some phonetic properties are spread over sequences of phones. These are called **prosodies** or **suprasegmentals**. Two prosodies are discussed in this section, pitch and stress. Others include loudness, tempo, length and rhythm.

Pitch

Pitch refers to the frequency of vibration of the vocal folds. When you speak the pitch varies from moment to moment. Variations in pitch are used in two main ways in languages: to distinguish between words; and to convey different inflections on the meaning of an utterance. In the former case, we speak of **tone**; in the latter, of **intonation**.

Tone

Many languages use different patterns of pitch to distinguish words; these pitch differences are called **tones**. Languages that use tones are called tone languages. Cantonese (Sino-Tibetan, China) is, like many nearby languages, a tone language. Differences in the pitch on the syllable [si] give six different words: with high falling tone, it is the word 'poem'; with mid-level, 'to try'; with low level, 'matter'; with extra low, 'time'; with high rising, 'to cause'; and with mid-rising, 'city'.[6] In fact, Cantonese has three more tones, making nine in total.

Intonation

All languages use variation in pitch over an utterance to convey modulations of the meaning expressed by the words. If you say *I'll see you tomorrow* as a plain statement, you will probably say it with fall of pitch at the end. If you want to ask your friend whether they will be coming into the university on the following day, you would normally utter it with a rise in pitch on the final word. Produced with a rising-falling intonation contour on *tomorrow*, the utterance would convey a degree of insistence. Pitch variations also convey other kinds of information, including information about grammatical structure and the speaker's emotional state – for example, whether they are angry, happy or sad.

The different intonation patterns and the variations of meaning that are expressed by them are difficult to specify exactly, and differ somewhat among dialects of English – as well as between

languages. For instance, in my dialect (Australian English) ordinary statements are often made on a rising intonation contour, which generally signifies a question in other dialects (see previous paragraph).

Stress

Syllables can be produced with different degrees of forcefulness or lung energy, normally accompanied by differences in the tension of the vocal folds. Increasing the energy gives greater intensity, loudness and usually higher pitch. Syllables with greater energy are called **stressed** syllables, indicated in the IPA by a ˈ before the syllable. Other phonetic differences sometimes correlate with stress. For instance, in English the vowel of an unstressed syllable is often schwa, as in the usual unstressed utterance of words such as *the* ([ðə]) and *a* ([ə]), and the second syllable of *farmer* ([ˈfɑːmə]). This vowel normally does not occur in stressed syllables.

In some languages stress always falls on a particular syllable of a word. In most Australian languages the first syllable of a word is always stressed. The following examples from Walmajarri (Pama-Nyungan, Australia) illustrate this: *ngarpu* [ˈŋaɻbu] 'father', *kurrapa* [ˈkuɾapa] 'hand' and *martuwarra* [ˈmaʈuˌwaɾa] 'river'. As the last example shows, the third syllable of a word with four syllables is also stressed, though usually less strongly than the first, indicated by the ˌ. Stress in Hungarian also goes to the first syllable of a word. In Swahili (Niger-Congo, Democratic Republic of Congo) and Polish (Indo-European, Poland), by contrast, it is the second-last syllable of a word that normally gets stress.

In English, stress is not predictable and goes on different syllables, depending on the word. Compare the placement in the three trisyllabic words (where the stressed syllable is bolded): ***pho****tograph*, *di****plo****ma* and *disa****gree***. There are a fair number of pairs of nouns and verbs in English that are distinguished by placement of stress. Stress goes on the first syllable of the noun, as in *an* ***im****port*, *a* **con***vict* and *an* ***in****sult*, but on the second syllable of the corresponding verb: *to im****port***, *to con****vict*** and *to in****sult***.

2.6 Phonology

How many phones does English have? Every time you utter a word or sentence there will be slight differences in the precise configuration of your vocal tract and the surrounding air. With sufficiently accurate instruments you could find minor differences in the shape of the vocal tract, and in the sound wave. (It's a bit like not being able to step into the same river twice!) In this sense you might say that English effectively has an unlimited number of phones. Many of the differences are too small to be perceived. Some differences are perceptible in principle – that is, are within the distinguishing capabilities of the human ear – but are ignored by speakers. **Phonology** investigates the sound differences that are linguistically relevant in a language, and how the sounds pattern as a system.