CHAPTER 1

The nuts and bolts

How speech works

Have you ever thought about your voice and the way you speak? Really thought about it?

If you're—for example—a singer, a spoken-word performer, have had speech and language therapy, or had an experience which has caused you to use your voice differently, I'll bet you have. But otherwise, unless you've been in a position where someone made some kind of comment about your voice or the way you speak, you may not have thought about it. Even then, you might have just ignored the comment as irrelevant, unimportant, or simply impertinent. But is it?

There is evidence to suggest that the way you speak is just as important now as it ever has been. All sorts of value judgments are made about people on the basis of their voice. The moment you open your mouth and utter that first phrase—no matter what you look like—conclusions are drawn; decisions are made. Some of those conclusions and decisions may have an effect on how successful you are in life.

This book is about the way that people speak and how their voice represents them, both how they perceive it to represent them and the perceptions of others. Among other things, it looks at what influences the way our mother tongue sounds when we

speak it, how we can classify and describe speech and accent features, whether the speech of people from different genders is similar or different, what people think accent tells us about speakers and the prejudices people have, how your voice can be used to identify you, how people modify the way they speak (or sing) depending on a variety of factors, and how advances in technology mean that individual voices can now be created for people who can no longer use their own.

The way people speak has always fascinated me, from the sound of the voice itself to the spoken words and grammar. This has probably got a lot to do with my father. Where he'd picked up his Southern British Standard (SBS) accent, I never knew . . . but he was certainly going to make sure my spoken English didn't descend too far into the local accent and dialect spoken where I grew up. Oh no—that would never do! And so I became fascinated by how people spoke, helped along by my mother's love of Hollywood musicals, and an introduction to the film *My Fair Lady*, which basically seemed to tell my father's story from the point of view of a London flower girl.

'I want to be a lady in a flow'r shop,' says Eliza Doolittle in George Bernard Shaw's play, ''stead of sellin' at the corner of Tottenham Court Road. But they won't take me unless I can talk more genteel.'

And what of Henry Higgins's (male, English-centric) assertion that 'An Englishman's way of speaking absolutely classifies him; the moment he talks, he makes another Englishman despise him'?

As a nation, the British seem to be much more accepting of regional variation than they used to be in, say, the 1950s. If you listen to BBC news and continuity presenters back then, they sound very upper class, usually having what is referred to as a

'cut-glass' accent, or 'The Queen's English'. The official term for the cut-glass accent is 'Received Pronunciation', or 'RP', where 'received' means 'accepted' (more on the subject of RP later). Until fairly recently, regional accents would only appear in fiction or drama on the BBC, not news and current affairs.

These days, there is a variety of accents on the air, and not just confined to regional news programmes. One might assume, therefore, that Henry Higgins's comment no longer applies in this day and age of social mobility. While I have heard tales that in the past BBC broadcasters had received letters complaining about their (usually very mild) regional accents, this seems to be less common nowadays, unless someone has quite a strong regional accent. Recently, BBC Breakfast presenter Steph McGovern, who hails from Middlesbrough and has a discernible Middlesbrough accent (but otherwise speaks with a Standard English dialect whilst on the air), disclosed that a viewer had sent her money for elocution lessons. Why should a regional accent be important when she has a degree in science and communication policy from University College London, over ten years' experience as a financial journalist, is a highly skilled communicator, and was named Young Engineer of the Year? Why should this educated woman's voice be the thing that defines her in the opinion of some members of the public? There has also been media speculation about whether Prince Harry's voice played a role in attracting Meghan Markle, whether footballer David Beckham ever had elocution lessons to sound more authoritative as England captain, and some rather unkind comments about actor and presenter Donna Air losing her Geordie accent to sound a little posher to date Kate Middleton's brother. Clearly, how you speak still matters to a large number of people, or it simply would not be newsworthy.

In this book, I'll be mainly using the word 'voice' in the non-technical sense, i.e., as a cover term for the way that people produce speech and how that sounds. In technical terms—as far as phoneticians and speech therapists are concerned, for example—'voice' refers to how lung air causes there to be vibration or friction of various kinds at the vocal folds in the larynx (sometimes known as the voice box), creating sound for humans to modify into speech and language. This sound may have a variety of voice qualities, depending on things like how the speaker feels, what emotions or effects they want to convey, or whether there has been damage to the vocal folds.

For example, think about your own voice when you're tired. You can actually hear the tiredness in the voice.

There will be some reference to this meaning of the word 'voice' when we look at voice quality later in the book.

As a starting point, we're going to look at how a child's voice is influenced by the speech environment in the early years. To quote Maria from *The Sound of Music*, let's start at the very beginning.

Babies, children, and fish

It is usual to think that a child's first experience of people's voices happens when they are born into the world and start to hear language spoken around them. Actually, this is not the case. A child's linguistic journey starts much sooner than that. It starts while the child is still in the womb.

It is worth pausing here to briefly examine the physical phenomenon of sound. We perceive sound by way of our hearing apparatus consisting of the outer ear, the middle ear, and the

inner ear. Sound itself comes to us from a source (for example, a person speaking, a musical instrument, the siren of an ambulance, the wind whistling down a chimney, the tinkling of a chandelier, the buzz of cicadas . . . the list is endless) and is carried through the air, moving away from the source in a series of waves which fan outwards.

Where speech is concerned, the speaker decides what it is they want to say, then the motor cortex in their brain sends signals to the muscles in the articulators, which include the lungs, the vocal folds (in the voice box or Adam's apple), the tongue, and the jaw. The lungs fill with air and the speaker starts to push the air out in a controlled way. The articulators move swiftly from position to position for various speech sounds, sometimes with the vocal folds producing voice, sometimes without, and these speech sounds are propelled via the pressure from the lungs into the air between the speaker and the listener.

Differences in air pressure due to the different kinds of sounds produced reach the ear of the listener. This causes the tympanic membrane, or eardrum, to vibrate, and three tiny bones (the ossicles) inside the inner ear transfer that vibration to liquid in the cochlea, or inner ear. The cochlea is shaped in a spiral, just like some seashells (for example, a conch shell or a whelk shell—you may have picked some up from the beach) and is filled with this liquid. Also inside the inner ear, spread along the entire length of the spiral, are tiny hairs, which vibrate according to the movement in the liquid caused by the movement of the three little bones. Electrical signals from the movement of these hairs then communicate to the brain what it is the listener has heard, and the brain finally makes sense of the message.

You can see a diagrammatical illustration of all of this in Figure 1.

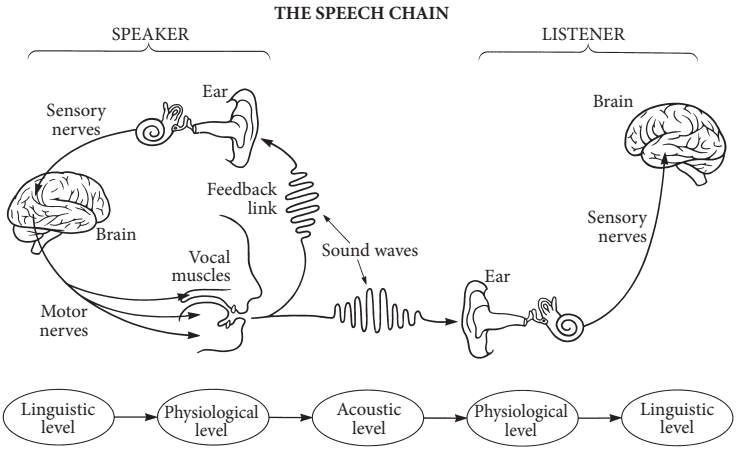

Figure 1 The speech chain

Air is important in transmission of sound from the source to your ear, as it acts as the carrier for sounds. There's a famous experiment with a ringing alarm clock inside a bell jar; when the air is pumped out of the bell jar, leaving a vacuum, it is no longer possible to hear the alarm clock ringing. As the tag line for the movie *Alien* observed, 'In space, no one can hear you scream'—that's because in space there is no air or other medium to carry sound.

Scan here or go to https://www.youtube.com/watch?v=hlOqX4uJtYY to see Tan Aik Hwee's YouTube video of the bell jar experiment

You might think that, if a speaker's mouth and listener's ears were not surrounded by air, it would not be possible to hear sounds. However, other media can transmit sounds, although the effect you get is different from when air is the medium of

transmission. Have you noticed when you're swimming that, to some extent, you can hear what's going on around you when you're underwater? And we get those hilarious cartoon-type high-pitched voices if you take in a mouthful of helium from a balloon[1]—but as soon as the helium disperses, we're back to boring old normal voice again.

Let's return to our unborn baby. A typically developing baby in the womb can hear the noises going on around their mother, even though the baby's ears are not surrounded by air. The baby can hear—and is listening to—all kinds of environmental sounds, the noises made by the mother's bodily functions . . . and speech.

All the parts of the inner ear described above are more or less in place by about the fifth month of a foetus's development. That means there is the possibility for sensitivity to sound from this point onwards. While we do not know exactly when an unborn child begins to attend to sound, what is clear from research carried out by Birgit Mampe, Kathleen Wermke, and colleagues[2] is that it is possible to identify which language group newly born babies (neonates) belong to based on the pitch patterns in their cries. The voice they will usually hear most loudly and clearly is their mother's and—while some of the sound is likely to come from the air around the mother—it will mainly be transmitted via the amniotic fluid around the foetus as the voice resonates through the mother's body. Try putting your fingers in your ears and count to five; you'll get an idea of what that might sound like.

The developing child will also be able to hear the voices of people their mother interacts with, the clearest signal being the falling and rising pitch of the voice. Again, if you think about how speech sounds underwater, we lose a lot of the fine detail but can still hear pitch differences. It's also a bit like when someone

has a loud party going on in the flat downstairs, when we might feel the sound more than hear it. Some people get quite excited around an expectant mother, and the physical manifestation of this includes increased pitch range; that is, their speech rises and falls in pitch in a much more exaggerated way. I've used capitals here for syllables which might have an exaggerated pitch: 'Oh my GOODness, you're PREGnant!' 'FanTASTic news!' 'When's it DUE?!' 'Is it KICKing? Can I FEEL??' etc. This is all being picked up and processed by the child's developing brain. I spent so much time around one of my friends when she was pregnant that, when the child was born, the only person other than his mother and father he would smile at for weeks following birth was me, and we think that's because he was already used to the sound of my voice and my pitch patterns.

The research showing that the cries of neonates have patterns of intonation similar to the language of their mother tongue is an indication that, as human beings, we are highly sensitive to speech. There are theories which claim we are cognitively predisposed to acquiring speech—in the 1960s, Noam Chomsky even suggested that we have a 'language acquisition device' in the brain (sometimes called the 'black box') which is only found in human beings. The existence of this 'black box' as a single, discrete area of the brain has not been proven; recent research using medical fMRI scans, for example, see that a number of areas across the brain are activated in speech and language acquisition and production. Researchers such as Sue Savage-Rumbaugh[3] and colleagues working with bonobo chimpanzees and other primates have also shown that it is possible for some non-human species to acquire and use language (but not speech) in a human-like way. What is certainly the case is that a newborn baby has physical, psychological, and social needs, and that

nature has seen to it that she is able to have those needs satisfied by effectively causing her to fit in immediately with the social and linguistic group into which she arrives; her cries sound like the speech of people from the community. Language is important from the moment the child is born.

What happens in the first five years of a child's life is the development of an adult-like linguistic system based on the languages to which the child is exposed. In my case, I was only exposed to English, with my main influence in early childhood being parents who only spoke with a fairly standard Southeast England accent. It's actually more common in the world to be exposed to more than one language from birth. This happens, for example, if a child's parents or carers are bilingual or multilingual, or speak different varieties of the same language.

I'm going to give a short summary here of the acquisition of English speech sounds, as speech is the focus of the book, but there are many other aspects of language acquisition. If you're interested in that subject area, I suggest having a look at Jean Aitchison's excellent book *The articulate mammal*,[4] which was first published in 1976 and is so popular that it is now into its fifth edition.

In 1960, child language researchers Jean Berko and Roger Brown documented something now known as the 'fis phenomenon'.[5] What happened was that a child referred to his toy fish as a *fis*, but refused to accept *fis* as the correct production of *fish* from an adult. The exchange went something like this:

Adult: What's that you're playing with?
Child: A fis.
Adult: Is that your fis?
Child: No. My fis.

Adult: Your fis?
Child: No! My fis! My fis!
Adult: Is that your fish?
Child: Yes! My fis!

Berko and Brown used this example to show that the child could understand adult phonology, and the difference between certain speech sounds, before he could produce all the sounds himself. This means that linguistic categories are developing in the child's mind faster than he can articulate them; he knows there is a difference between the sounds at the end of *fis* (the phonetic symbol for it is [s]) and *fish* (phonetic symbol [ʃ]) but does not yet have enough control over the articulators in his mouth (mainly the tongue here) to make that difference. People have replicated this with their own children and posted videos on YouTube; search for 'fis phenomenon'.

We can tell that the child in Berko and Brown's study was probably at least three years of age, but he could be as old as seven. How do we know that?

In my work with speech and language therapy students, we have to consider what typically developing children do when they are acquiring their first language in order to work out whether a particular child has a problem—a speech and/or language deficit. One of the things we refer to is research on when children are able to produce certain kinds of speech sounds and use them correctly. Various studies have identified that there is an order of acquisition of consonant sounds in English, as there is in all languages. The order is shown in Figure 2, which indicates the age at which half of all typically developing children produce a consonant and use it correctly.

You can see that there is a lot of variation possible; a child might take up to five years to be able to produce, for

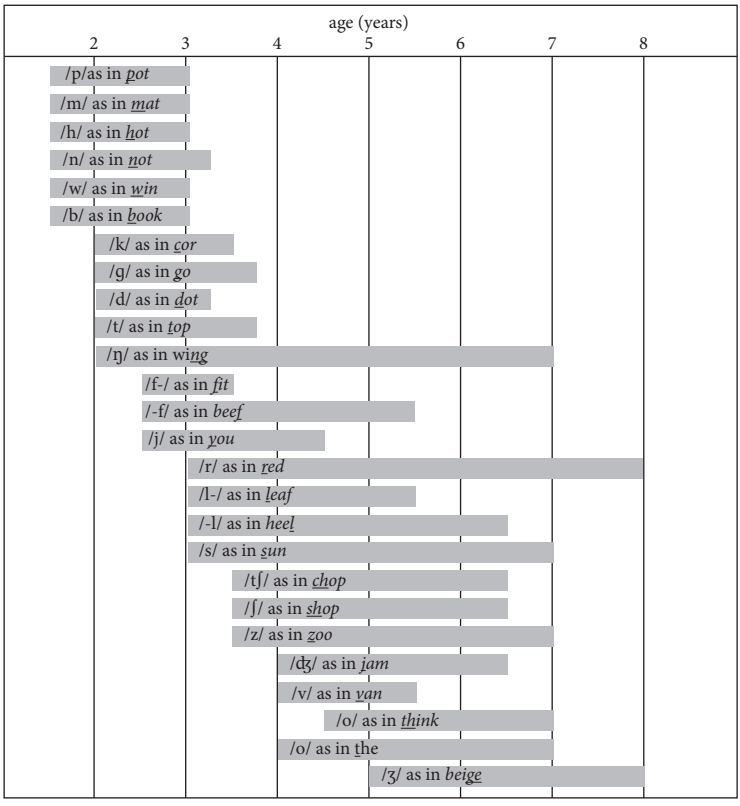

Figure 2 Age of acquisition of English speech sounds

example, /ŋ/ sounds (spelled *ng*) and /r/ sounds in an adult-like way, whereas the sounds /p t k b d g m n w/ and /h/ are usually acquired by the age of three-and-a-half. In the example from Berko and Brown, we can see that the child can produce /s/ but not /ʃ/, and so must be in a transitional stage between the ages of three and seven. We would need other data—including vocabulary scores and grammatical features—to be able to work out more precisely how old the child is.

Let's take some hypothetical data based on real speech from a child of 2.5 years, Nellie, interacting with her mother, Dawn. Dawn tries to get Nellie to say the word *fish*, and Nellie's production is variably something like /khap/, /jap/, and /hap/. This tallies well with the information in Figure 2. We would not expect Nellie to be able to produce the /f/ or /ʃ/ sounds in *fish* at the age of two-and-a-half, and indeed she does not produce them but uses other sounds fairly variably, with the exception of /p/ at the end of the word. But when Dawn repeats exactly what Nellie said back to her, Nellie rejects it until her mother correctly says the word *fish*. That's the 'fis phenomenon' at work. Babies and children are not just hearing language . . . they are listening and learning.

Although there are many differences in personality and how people's brains are wired, humans are basically social beings, and this also has an effect on language from an early age. As far as accent features are concerned, children's voices will usually sound very much like their caregivers' voices in the early years. Over time, as they come into contact with other people (for example, childminders and teachers) and get involved with different social groups (friendship groups, clubs and societies, even people from the television), the way they speak is likely to change. This change in the way people use their voices and adapt their accents according to different social situations continues throughout life for many. Do you have a 'telephone voice' when you have to call someone you don't know socially? Does your teenage son or daughter speak differently with you than they do with their best friend, or the parents of their best friend? If they do, this is normal, and the speakers themselves probably don't even realize it's happening (although I have friends who admit to having a telephone voice). Our voices

evolve to show our social allegiances, our tribal memberships. This chameleon-like aspect of your voice is called code-switching or accommodation, and we are all speakers of many different codes.

Sounds, patterns, and codes

Understanding and producing language is all about being able to decipher and apply patterns and codes. Where speech sounds are concerned, our brains are basically biological computers performing huge feats of code-breaking every moment of every day. We need to be able to identify a given speech sound when a speaker produces it in the context of other speech sounds as part of a linguistic message. Not only do different people from the same accent group produce speech sounds slightly differently from others, every time we utter a speech sound, it is different. Amazingly, our brains are able to work it out. This can be more of a challenge when faced with an accent we don't understand but, given experience with that accent, our brain works it out.

So, what are the speech sounds of English?

There are many varieties of English (and many languages) in the world, so I'm going to stick to one British accent which is well documented as a reference from which to describe others. This accent is variously known as BBC English, General British (GB), Received Pronunciation (RP), or Southern British Standard (SBS), and is the variety usually found in dictionaries, reference books and British English language teaching materials.

This does not mean it is the 'correct' accent for English. It is one of many.

I'm going to follow Alan Cruttenden's terminology in the most recent edition of *Gimson's pronunciation of English*[6] and use General British in the sections which follow.

Consonants

GB is usually described as having forty-four individual phonemes (for the purposes of this description, a phoneme is a speech sound in a given language, but there will be some more discussion of this term later on). Of these forty-four sounds, twenty are vowel sounds and twenty-four are consonants. The *Cambridge English pronouncing dictionary*[7] gives the following keywords for GB consonants, with all but one (*hang*) referring to the first consonant sound in the word:

Voiceless	*Voiced*
p as in *pea*	b as in *bee*
t as in *tea*	d as in *do*
k as in *key*	g as in *go*
f as in *fat*	v as in *vat*
θ as in *thin*	ð as in *that*
s as in *sip*	ʃ as in *ship*
h as in *hat*	
	m as in *map*
	n as in *nap*
	ŋ as in *hang*
	l as in *led*
	r as in *red*
	j as in *yet*
	w as in *wet*
tʃ as in *chin*	dʒ as in *gin*

You'll have noticed these are organized with some on the left and some on the right. The symbols on the left represent consonant sounds produced without vibration of the vocal folds ('voiceless'

consonants), and the ones on the right are produced with vibration of the vocal folds ('voiced' consonants).

To test this, put your hand on your throat where your voice box is, and say the sound /s/ as long as possible: 'sssssssssssss'. You shouldn't feel any vibration in the larynx when you say this sound.

Now change to /z/: 'zzzzzzzzzzzzzz'. You should be able to feel vibration when you say this sound. (I love doing this with my students and seeing their reactions.)

This technique works well for all consonant sounds except the six at the top of the list, voiceless /p t k/ and voiced /b d g/, and the two at the bottom, voiceless /tʃ/ and voiced /dʒ/. This is because of the type of sounds they are. With the other sounds, as long as you have air in your lungs, you can keep producing them. But these eight sounds involve stopping the air and then releasing it in a burst.

Let's try saying the sound /p/ as an example. If you put your hand in front of your mouth and say 'aapaa', you should be able to feel the warmth of your breath for the first 'aa', then everything stopping as you close your lips to say the /p/, and then a burst of air as you open your lips and resume the 'aa'. These sounds are called 'plosive' consonants because there is an explosion of air as you say them. We get the explosion because the air is held in the mouth and can't go anywhere else, causing air pressure to build up until the closure (in this case, the upper and lower lip pressed together tightly) is opened.

Because you have to stop the air to say them, it's very difficult to compare the sounds /p/ and /b/ in the same way as, for example, /s/ and /z/. So, how can you compare them and tell them apart? By comparing the strength of the burst of air.

If you put your hand in front of your mouth and say 'aapaa' and 'aabaa', you should be able to feel that the /p/ sound has a

stronger burst than the /b/ sound. This difference is in something known as aspiration; in GB, we say that /p/ is aspirated (has a strong burst of air after it) and /b/ is not aspirated. However, accents of English differ in this respect, so you may not have the same pattern. I demonstrated this phenomenon on the Alan Titchmarsh television show, and we discovered that Alan's accent might have different patterns of aspiration.[8]

/tʃ/ and /dʒ/ in *chain* and *Jane* are not plosives but affricates. An affricate is a sound which begins by stopping the air, like a plosive, but—instead of releasing the air relatively quickly and cleanly—the air is released in the manner of a fricative sound. Compare *tip* and *chip* which start with /t/ and /tʃ/ respectively.

Which sounds are fricatives in English? /f θ s ʃ h/ (voiceless) and /v ð z ʒ/ (voiced).

I've explained how plosives are produced. How about fricatives?

Fricatives are like plosive sounds in that they are produced under pressure. A small gap is made in the articulators and the air is forced through the gap. This results in a hissing sound. The hissing is most obvious with /f θ s ʃ/. Try producing each one of those on their own and observe the hissing noise.

Plosives, fricatives, and affricates all involve an obstruction to the air, in which the air is under pressure in the mouth, or oral cavity. In order to achieve this pressure, one other vital component is needed: the air must not be allowed to pass through the nose. This means the soft palate must be raised to block the air off. I think it's time to have a look at the articulators.

The picture in Figure 3 is a diagram of the side of the head, as if someone had sliced from top to bottom. It is labelled with terms describing parts of the oral cavity (inside of the mouth) involved in the articulation of speech. What we can't see clearly here are

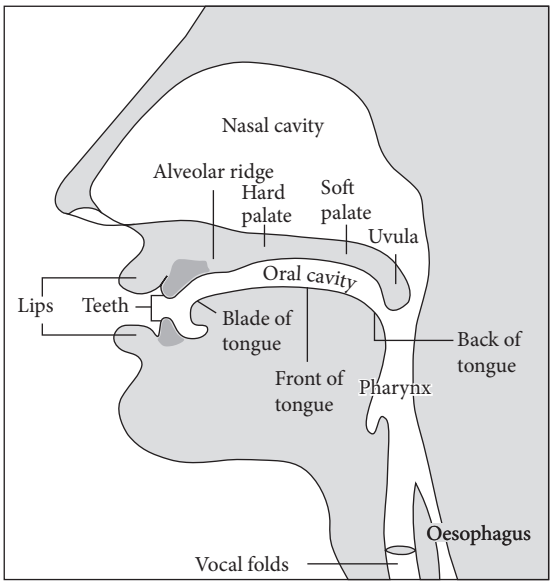

Figure 3 The organs of speech

the sides of the tongue, or the teeth, which go all the way around on both sides of the oral cavity.

Let's take a journey through the articulators. Start by pressing your lips together. In the articulation of speech sounds, we talk about active and passive articulators, with the active articulator moving towards and/or coming into contact with the passive articulator. It's usual to think of the parts labelled across the top of this diagram as passive, as they can't move, and the lower lip and tongue to as active, as they are more mobile and move around easily. We can, of course, move our top lip quite a lot (try it!), but it is fixed to the upper mandible (just above your top teeth and below your nose), which doesn't move. Your lower lip is attached to the lower mandible, which is part of the jaw and highly mobile. You can press your lips together without moving

the jaw very much, but if you open the jaw your bottom lip will move away from the top lip. Sounds made using both lips pressed together are called 'bilabial' sounds (literally, 'two lips'), and in English these are /p/, /b/, and /m/.

Another sound we make using the lips is /w/, which involves rounding the lips and pushing them out slightly. We'll come back to /w/ later as there's more going on than just lip movement.

We can also make sounds with the bottom lip against the upper teeth. These sounds are 'labiodental', and in English are /f/ and /v/. Try making these sounds.

Now I want you to run your tongue tip slowly from the cutting edge of your teeth back towards the velum, also known as the soft palate. As you run it backwards, you can feel: 1) the backs of the teeth; 2) a bony ridge, followed by 3), a sudden upwards curve; 4) the curve of the roof of your mouth, first bony, then 5) soft. We make dental sounds /θ/ and /ð/ between the back of the teeth and the tongue tip. The bony ridge is called the alveolar ridge, and sounds produced there are alveolar consonants /t/, /d/, /s/, /z/, /n/, and /l/. Say the words *tip*, *dip*, *sip*, *zip*, *nip*, and *lip*, which begin with these consonants, to try them. Some types of /r/ sound are also made there. Can you roll your /r/ sounds using your tongue tip, like in Spanish? That's made on the alveolar ridge. My primary school singing teacher always got us lot in the choir to roll our *r* sounds. The technical term for this is 'trilling'.

Sounds /ʃ/, /ʒ/, /tʃ/, /dʒ/, and what's thought of as the 'normal' General American /r/ sound are made with the tongue behind the alveolar ridge, at the sudden curve upwards—number 3) above. For each of these sounds, say the words *shoe*, *jus* ('zhoo'—the French word used in cooking—OK, I know it's a bit pretentious, but /ʒ/ is a pretty infrequent sound in English!), *chew*, *Jew*, and *rue*, which all rhyme, and focus on the initial

consonant sound. If you compare *sip* with *ship*, you should be able to feel the difference in tongue position between alveolar sounds (/s/ is alveolar) and these ones, which are called post-alveolar (that is, just after the alveolar ridge).

The next sound to try is /j/ at the start of *yet*, which is palatal. Don't confuse it with the letter *j*, which is usually produced as the affricate /dʒ/ as in *Jane*. When you make the /j/ sound, the tongue bunches up and approaches the hard palate but doesn't actually touch it (but you should be able to feel the sides of the tongue against the teeth). This type of sound is called an approximant, and involves bending the air around rather than stopping it or forcing it through a narrow gap. Another type of approximant is a lateral approximant, and in English the only one is alveolar /l/. We call this lateral as we make tongue-tip contact on the bony ridge behind the teeth, but the sides of the tongue are low. If you make the shape for /l/ and breathe in, you should be able to feel the cold air flow along the sides of your tongue.

The last approximant sound to try is /w/. We already met this sound when we were looking at lip movement, but there's more to /w/ than meets the eye—quite literally. We can see the lips rounding when someone produces /w/, but something else is going on; the back of the tongue moves up towards the soft palate at the same time as the lips are rounded, and this is not visible from outside. Try saying 'aawaa' and focus on what is happening with your tongue. You should feel the back bunches up a bit and moves upwards. We call /w/ 'labial-velar' as it has two places of articulation: the lips round, and the back of the tongue moves towards the soft palate.

Other sounds involving the back of the tongue and the soft palate are /k/ in *core*, /g/ in *gore*, and /ŋ/ at the end of *song*. /k/ and /g/ are both types of plosive, completely blocking the air, and

/ŋ/ is a nasal consonant. Try saying them and feel where your tongue is as you do.

/m/, /n/, and /ŋ/ are all nasal consonants. If you've got a cold, you may not be able to produce these sounds or breathe in on them. That's because, when we make these sounds, the air escapes through the nose. If your nose is blocked, an /m/ will sound like a /b/, a /n/ like a /d/, and a /ŋ/ like a /g/. We'll see in Chapter 4 how one particular type of voice uses that to advantage.

Nasal consonants are exactly the same as plosives, except that the soft palate is lowered so that air pressure does not build up in the oral cavity but passes out through the nose. In the case of /b/ and /m/, for example, the soft palate is raised for /b/—as we don't want the air to be able to escape until we open the lips—but lowered for /m/ so the air can pass out through the nose. When we produce /b/, there is air pressure build-up, but not for /m/.

You can feel this physically by trying to say this sequence of sounds without opening your lips: /bmbmbmbmbmbm/.

Can you feel the differences in air pressure?

You may also be able to feel your soft palate lowering in the upper back region of your mouth as you do this. We don't usually think about what the soft palate is doing and so don't generally notice it.

Another way we can tell the air is coming through the nasal cavity is to place the glass of your mobile phone or a small mirror under your nose and then produce two sounds, one without nasal airflow and one with.

First, say a long, sustained /s/.

/sssssssssssss/

What can you see on the mirror / your phone? (Tip: there shouldn't be anything there as /s/ does not have nasal airflow.)

Now try a long /n/.

/nnnnnnnnnn/

You should be able to see water vapour on the glass from the air coming up from your lungs and through the nasal cavity.

One consonant which is not on this list but is very common in English accents is called a 'glottal stop' or 'glottal plosive', and it's made by pressing the vocal folds tightly together. The symbol for this sound is [ʔ] . . . and you may notice it's in a different set of brackets from the others. There's a reason for that.

First, try making this sound. You can do this by coughing very gently. When you do this, you close the vocal folds, hold them shut for a short while, then release them to produce the burst of air we associate with all plosive consonants. We cough because we are trying to remove an irritant in the throat, and the only way the body has of doing this is using highly pressurized air. The best way of producing air pressure build-up in that area is to close the vocal folds and push air out of the lungs.

Now try producing this sound as a speech sound. You may have it in your accent anyway. Say the word *hotter*, and try doing the gentle cough instead of the /t/ sound. If you do this, both the spoken consonants in this word in the GB accent are glottal—that's the /h/ and the [ʔ].

The glottal stop gets a bit of a bad press. It's particularly associated with London accents spoken by people from less educated backgrounds. When speakers use it to replace /t/, they are often described as speaking in a slovenly manner. But actually, it's used by speakers from many backgrounds, and is not out of place in the reference accent, GB. The main difference is that accents considered to be less prestigious use it in words like *hotter* where it is between two vowel sounds. In more prestigious accents, it tends to get used more often

at the ends of words and not between vowels, and is therefore less obvious.

Take the phrase 'isn't it?' Where a speaker of old-fashioned RP would most likely pronounce both /t/ sounds, a modern speaker of GB is more likely to have a glottal stop at the end of *it*.

But why have I put it in square brackets, [ʔ]? This is because, in English, it always stands in for another plosive consonant—usually /t/. When a speaker uses it and a listener hears it, they understand /t/. If I substitute [ʔ] for /t/ in a word like *hotter*, it makes no difference in meaning. [ʔ] is therefore a production variant of /t/. We call this an allophone.

However, if I substitute any of the other sounds for /t/, potentially I can change the meaning of the word. All of the words *sat*, *sad*, *sam*, *sang*, *sap*, *sack*, *sag* start with the same two sounds, but the final consonant is different in each case, and they all mean different things. This shows that /t d m ŋ p k/ and /g/ are all different meaning units in English: they are phonemes. If I have [ʔ] at the end, however, listeners will perceive *sat*. It doesn't mean anything different from /t/.

To use a phrase from *The Matrix*, this might be baking your noodle right now. Let's move on to vowels.

Vowels

Vowels have got to be less complex, right? After all, English only has five of them: A, E, I, O, and U . . . doesn't it?

Nope.

The General British accent actually has twenty vowel phonemes, and a couple of others thrown in for good measure. It's complex. Anyone working with children who are learning to spell will realize this is not a straightforward area.

For vowels, I will use the keywords from John Wells's Standard Lexical Sets.[9] One of the reasons for doing this is because almost all work describing accents of English since the publication of the Standard Lexical Sets in the 1980s uses the Wells framework to talk about vowel variation.

The words to represent each vowel sound given in CAPITALS below were chosen because they cannot easily be confused with other words. The remaining words in each line are examples of how the vowel appears in English spelling.

1. KIT ship, kid, limp, myth, build . . .
2. DRESS step, ebb, tent, bread, friend . . .
3. TRAP tap, rag, hand, lapse, plaid . . .
4. LOT stop, odd, box, swan, wash . . .
5. STRUT cup, bud, lump, come, touch . . .
6. FOOT put, bush, good, wolf, could . . .
7. BATH staff, class, ask, fasten, laugh . . .
8. CLOTH off, cross, soft, cough, Austin . . .
9. NURSE hurt, birth, church, verb, word . . .
10. FLEECE creep, need, cheese, brief, field . . .
11. FACE tape, fade, waist, play, reign . . .
12. PALM calm, ma, hurrah, façade, Java . . .
13. THOUGHT cause, taunt, hawk, chalk, broad . . .
14. GOAT soap, joke, host, toe, mauve . . .
15. GOOSE loop, mood, tomb, two, fruit . . .
16. PRICE ripe, side, child, try, eye . . .
17. CHOICE boy, noise, spoil, employ, hoist . . .
18. MOUTH out, crowd, cow, round, bough . . .
19. NEAR beer, here, pier, fear, pierce . . .
20. SQUARE share, fair, bear, where, scarce . . .
21. START far, sharp, carve, heart, safari . . .

22. NORTH for, orb, form, quart, cord . . .
23. FORCE fore, soar, floor, court, sword . . .
24. CURE moor, your, sure, gourd, fury . . .
25. happY copy, khaki, movie, coffee, money . . .
26. lettER paper, sugar, standard, anchor, martyr . . .
27. commA quota, visa, panda, sofa, saga . . .
28. thankYOU evaluate . . .

As you're reading through these words, you might be thinking to yourself, 'Hang on—I've got the same vowel sound in TRAP and BATH', or 'The way I say NORTH is no different from how I say THOUGHT or FORCE' (or even CURE). That's the clever bit about this list. John Wells noticed that there are groups of words which some people say differently and some do not, and it is from these observations that this list arose. The differences are mainly down to the historical development of English accents over time in different areas of the British Isles, in other countries where English is spoken, and also owing to differences in social status. Most people from the North of England and from the United States, for example, have the same vowel in TRAP and BATH, but for the majority of speakers from the South of England the vowels will be different. What should not be the case is that you say words in each set differently; for example, all words in the START set should have the same vowel, whether or not START has the same vowel as BATH.

Here are the words in the lexical sets organized according to the vowel sounds they have in General British, together with the vowel symbols used in this book; these are the ones you see in most British English pronunciation text books. Both northern and southern standard variants are indicated, where relevant.

'Short' vowel monophthongs (six of these)
KIT /ɪ/
DRESS /e/
TRAP /æ/
LOT, CLOTH /ɒ/
STRUT south /ʌ/, north /ʊ/
FOOT /ʊ/

'Long' vowel monophthongs (five of these)
BATH south /ɑ:/, north /æ/ (short in northern accents);
PALM, START /ɑ:/
NURSE /ɜ:/
FLEECE /i:/
THOUGHT, NORTH, FORCE /ɔ:/
GOOSE /u:/

'Long' vowel diphthongs (eight of these)
GOAT /əʊ/
MOUTH /aʊ/
FACE /eɪ/
PRICE /aɪ/
CHOICE /ɔɪ/
NEAR /ɪə/
SQUARE /eə/—sometimes transcribed /ɛ:/
CURE /ʊə/

Weak vowels (three of these)
lettER, commA /ə/
happy /i/
thank yOU /u/

Let's look a bit more at the set described as 'weak vowels'.

Traditionally, the vowel at the end of lettER and commA, known as 'schwa' and symbolized /ə/, is included in the twenty vowels of General British . . . but the happY and thank yOU vowels are not. This is because, among speakers with older versions of RP, the happY and thankyOU vowels would have been pronounced /ɪ/ as in KIT (for happY) and /ʊ/ as in FOOT (for thankyOU). Speakers of General British tend to pronounce the vowel at the end of happY as more of a cross between the FLEECE vowel and the KIT vowel, and the one in thank yOU as a cross between the GOOSE and the FOOT vowel. The important thing to note is that we are only talking about these vowels in certain positions, which is that they are never stressed.

Something else you may have noticed is that vowels are described as monophthongs and diphthongs. What's the difference between a monophthong and a diphthong?

When we produce vowel sounds, the vowel we get depends on where our tongue is positioned in the mouth. The vowel in START (in British English) is very like the one doctors ask you to say when they want to look at the back of your throat—'Say "aaaahhhh", please!' (depresses tongue with flat implement). Try saying 'aaaahhhh' now. Your tongue stays in one place when you say this vowel, and so we describe it as a monophthong ('mono' = one; 'phthong' = sound).

If you say the vowel in PRICE, however, your tongue starts in one place and moves to another. It goes on a little journey. We can see and feel this better with the word *eye*, which contains only that vowel with no consonant sounds. Try it. Can you feel your tongue and jaw closing as you produce this vowel? As there are two tongue positions involved, we call it a diphthong ('di' = two; 'phthong' = sound).

The terms 'long' and 'short' are also used in the list. The difference between long and short vowels depends on context and position in a word. The position part can be dealt with the most easily: there are no single syllable content[10] words in GB containing a short vowel which do not end with at least one consonant, either in spelling or in pronunciation. Examples include *cat*, *sing*, *shot*, *hut*, and *fed*. Single- syllable words containing long vowel monophthongs and diphthongs, by comparison, do not have to have consonant sounds after them, but they can do. Examples without consonant sounds include *play*, *lie*, *boy*, *see*, and *saw*.

When the Ford motor company decided to name its smallest car the KA, this drove me nuts. Why? Because of the rule above, people felt compelled to pronounce KA as *car* (/kɑː/), containing the START vowel, as it cannot be a short vowel. But to me—from both a spelling and pronunciation point of view—it was /kæ/, with a TRAP vowel . . . which is illegal in GB. I'm told by people in car sales that it's supposed to be pronounced /keɪ eɪ/ ('kay-eh'), but I've never heard anyone call it that. Have you read about those people who wander round with a marker pen correcting bad grammar on signs? I think Lynne Truss says something about this in her book *Eat, shoots and leaves.*[11] I've managed to stop myself adding the letter *r* to the logo on Ford KAs, but it makes me cringe every time I pass one. If it's supposed to be pronounced 'kay-ey', it should be K.A., surely?

Stress and intonation

When I was talking about babies earlier in this chapter, I mentioned that the pitch of the voice rises and falls as people speak, and that this is increased when they are excited about something—that is, when they are emotionally involved.

Intonation is sometimes described as 'the melody of speech'. It is very unusual in human languages for speech to be produced without melodies, or to be monotone (like the speech of the Daleks in *Doctor Who*—but even these sometimes have a rise in pitch at the end). As with all linguistic features of speech, these melodies have patterns which are specific to particular languages.

In some languages—like Chinese dialects and indigenous languages spoken in Africa and South America—changes of pitch on a single syllable can change the meaning of the word. These are called tone languages. A commonly given example is how tones change the meaning of the syllable *ma* in Mandarin Chinese. Mandarin Chinese has four tones and, depending on which tone you use, this syllable can mean 'mother', 'horse', 'hemp', and 'to scold'. (You can easily find examples of this if you do an online search.)

English does not operate in the same way, but we do have some interesting patterns of word stress and intonation which can change the meaning of a word or an utterance.

There are several words which change from nouns or adjectives to verbs if the stress is different. For example, if the word *record* is a noun (meaning item on or in which information is stored), we stress the first syllable—i.e., the first syllable is louder and longer than the second syllable and the pitch is slightly higher: *RECord*. If it's a verb (meaning to create that item on or in which information is stored), the stress is on the second syllable: *reCORD*.

Think about the word *perfect*: what does it mean if you pronounce it *PERfect* in comparison with *perFECT*? ('That's the perfect gift for her,' versus 'I'm going to perfect my icing technique.')

Intonation is speaker-specific as well as language specific. What I mean by this is that everyone has their own average

spoken pitch and pitch range. We measure spoken pitch in Hertz, which is basically the number of vibrations per second of the vocal folds during voicing. If you growl in your throat, you may be able to hear individual vibrations of the vocal folds (this can be used in speech and is referred to as 'creak' or 'vocal fry'—more on this in Chapter 3). In normal speech, the vocal folds usually vibrate much more quickly than this. Among men, average pitch is around 100–150 cycles (vocal fold vibrations) per second. Among women, it is usually double that, and among children it is three to four times more. Average pitch usually drops as people age—very noticeably so for men, whose voice 'breaks' during puberty and suddenly becomes much lower in pitch.

In languages, many features of intonation are similar. For example, if you ask a question, your pitch often goes up at the end (but not always!) and, in statements, it often goes down. This is common across many languages. However, as we'll see in the chapters which follow, this is not always the case, and there are some interesting patterns of pitch and intonation use which can signal which social group an individual belongs to, as well as indicating linguistic meaning.

Patterns and codes

Every language has its own set of distinctive sounds, and these sounds operate in patterns, such as the rules in the GB accent where the glottal stop can only stand in for /t/ when it is not between vowel sounds, or /ŋ/ only occurring at the ends of words. In Cantonese, for example, /ŋ/ can occur at the beginnings of words—the word for *I* in Cantonese is /ŋo/, spoken with a rising tone. As speakers of particular languages, these rules form during childhood and can be hard to shift in adulthood,

which is one reason why it is more difficult for us to learn languages later in life, and particularly difficult for language learners to have native-like accents.

This is also the case for the patterns in different accents of a language. Just because someone speaks a given language (e.g. English), it doesn't mean their accent will be easily understood by other speakers of that language, as the patterns can be very different. We'll look more at this in Chapter 2, where we look at British English accents and accent prejudice.

What is certain, however, is that languages are rule-based and have identifiable, regular patterns. If they were random, we wouldn't be able to understand each other. We rely on patterns in order to be able to decode the message we are listening to. Every speaker has slightly different patterns, but they correspond generally to the overall pattern of a particular accent. This is why, when you hear someone you haven't met before from the same accent group as you speak for the first time, it can take a couple of seconds to understand them, but our brain is really good at accommodating to new voices if the accent is one you are familiar with. It's also one of the reasons we have predictable social greetings like 'Nice to meet you', as it gives the brain time off from understanding the message so it can calibrate itself to the speaker's voice and accent. If someone says something less predictable, it can throw the brain off. And if someone has a less familiar accent, it can take a lot longer to get used to the patterns.

Here's an example from when I was a child. I had a school friend whose parents were from Scotland, and her father was from Glasgow. When I met him, the first thing he said wasn't something predictable like 'Hello' or 'Nice to meet you', which are quite common as greetings. Oh no. It was 'Will you no have a chair?'

And I didn't understand him.

Not just because the grammar was unusual to me, or because it was an unexpected greeting, but also because his accent and pronunciation were very different from anything I had ever heard.

My friend's mother—whom I'd met before and whose accent was not so broad—then said, 'Won't you sit down?' It was only at that point that her father's words suddenly made sense. My brain needed time (and a translation) to work it out.

After that, I had less difficulty understanding my friend's father, as I'd got used to his voice . . . but it did take me a bit longer to get used to his accent.

What's coming up?

Here's an overview of the rest of this book.

In Chapter 2, I'll be looking at accents of the UK. The focus here is on how speakers of different accents are perceived, but there will also be an overview of how we came to have all these accents in the first place. You may be surprised to learn just how hybrid and 'foreign' the English language is.

There are lots of myths about speech and language, and Chapter 3 covers those specific to speech. Can men make their voices sound sexy? Do only young people have the kind of intonation which goes up at the end? Are women's voices 'worse' than men's voices? And so on.

Had you noticed how some singers have one accent when they sing, and another when they speak? Chapter 4 investigates this phenomenon, and also looks at other language professionals and whether they change the way they speak in their professional environment.

Can speech be used to catch criminals? In my role as a forensic phonetician, I have worked on a number of cases comparing the speech from, for example, telephone recordings with police interviews. Chapter 5 explains how this works, and also considers the field of speaker profiling, in which phoneticians try to identify a person's origins from a speech sample.

Chapter 6 is on people whose voices have been changed in some way other than accent features to reflect their identity. There are examples from transgender speakers, and machines which produce speech for people who are unable to do so themselves.

Finally, Chapter 7 rounds off the discussion by looking briefly at speakers of English around the world, including some observations on speech and ethnicity. There is then an annotated bibliography (list of books and papers) for anyone wishing to delve into the subject of voice in any greater detail.

I hope you enjoy the book, and find this subject as fascinating as I do. If you'd like to chat with me about it, or even take me to task, you can find me on Twitter: @JaneSetter.