2
Making speech sounds

Speech and breathing

In the languages of the world there is a huge and fascinating variety of different speech sounds, and we need to understand how these speech sounds are made. The most basic fact to remember is that all the sounds we use in speech are produced by moving air. Speech is sometimes described as 'modified breathing': the process of speaking begins, except for a relatively small number of cases, with the air inside the speaker's chest escaping from the lungs, through the throat and the mouth, and out into the open air, as you can see in Figure 2.1. If we produce this flow of air without impeding it in any way, the activity would just be what we call 'breathing out', or, if we do it loudly enough to make a sound, a 'sigh' (which can convey many different meanings). But usually in speaking we use our **articulators** to modify the flow of air so that sounds are produced. When we do this, we produce a sequence of vowels and consonants that make up **syllables**: if you say 'Cut it out', for example, you are alternately obstructing the flow of air from the lungs (when you produce the consonants) and then allowing the air to escape rather easily (in the vowels). In this example, three syllables are produced. Syllables usually contain a vowel, and may start and end with one or more consonants.

It is a very interesting fact about speech that it has evolved by making use of parts of the body which already have some biological reason for being there, and as far as we know there is nothing in the human body which exists exclusively for making or recognizing speech sounds. We would still need our lungs, our tongues, our vocal folds, and our ears, even if they were not required for speech.

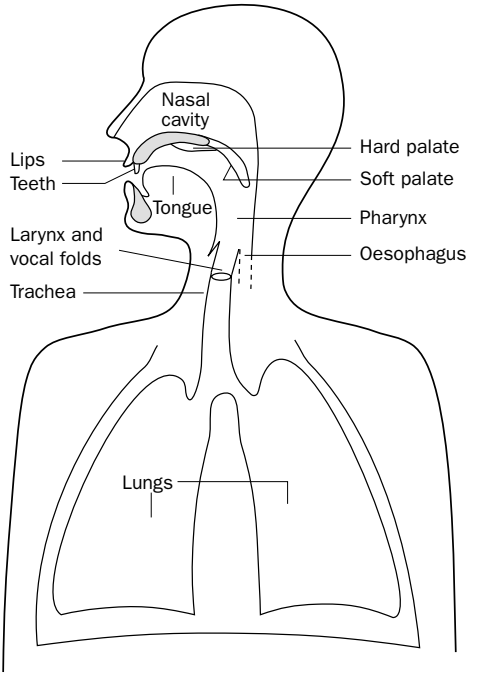

FIGURE 2.1 *The breathing system and the vocal tract*

There are, in fact, some speech sounds which are made by using something other than the lungs to make the air move. To English speakers, many of these sounds are familiar but 'non-linguistic'—they are not used as phonemes of the language: for example 'click' sounds are found when we make the 'annoyance' noise that used to be written as 'tut-tut' (or nowadays more usually as 'tsk-tsk'), and people still sometimes use a 'gee-up' click to tell a horse to move on. But in a number of languages of southern Africa such as Zulu and Xhosa, we find a wide variety of click sounds being used as consonant phonemes of the language. There are also sounds known as 'ejectives' and 'implosives' in many of the world's languages, and to produce these you have to use the larynx: the vocal folds are closed or nearly closed, and the larynx is moved upwards (in ejectives) to push air out, or downwards (in implosives) to suck air in. These sounds are found in many

different languages, including some accents of English. Amharic (spoken in Ethiopia) is a good example of a language with ejective sounds, having contrasting pairs of phonemes /t/ and /tʼ/. The latter is ejective: the way ejectives are produced is a little like the mechanism used for spitting some small object (e.g. a grape pip) from your mouth. An example of a language with implosives is Sindhi (spoken in India). This has a /b/ similar to that of English and a different phoneme /ɓ/ in which air is 'gulped' inwards by a movement of the larynx.

The passageway through which air passes from the larynx, past the lips, and out into the air outside our bodies is called the **vocal tract**. Below this is the **trachea**, the 'windpipe' which is connected to the lungs and which passes from the chest into the neck. We do not have the ability to move or modify this.

The larynx

The **larynx**, indicated in Figure 2.1, is very important. The vital part of the larynx is a pair of folds of muscular tissue called the **vocal folds**, and we can move these into a number of positions between wide open and tightly closed. We open them widely to allow a rapid escape of air. If they are slightly narrowed, so that the gap between them is only a few millimetres, the air makes a rushing noise that we associate with the sound at the beginning of the English word 'head'. If we close them enough for them to be lightly touching each other, the air passing between them causes them to vibrate; this is called **voicing**, or **phonation**, and it can be varied in many ways, most importantly in **pitch**, which may be high or low according to how we adjust our vocal folds. Many speech sounds are **voiced**, while others are **voiceless**. If you want to practise detecting the difference, compare [s] and [z]: the consonant [s] is voiceless and the only sound it makes is the hissing sound as air escapes over the tongue. However, its voiced counterpart [z] not only has this hissing sound, but also the buzzing of the vocal folds vibrating. You can hear this best if you put your fingers in your ears and produce a long, buzzing [z]. You can also detect the vibration if you press your fingers gently against your larynx as you produce the [z]. Vowels are almost always voiced, as are **nasal** consonants like [m] and [n]. Finally, if

we close the vocal folds firmly, we prevent air from escaping. This is often called a **glottal stop**.

The vocal tract above the larynx

Immediately above the larynx is a passageway called the **pharynx**. This carries air, but also food when we are eating. As we all occasionally find when we are trying to eat and speak at the same time, this passageway can get rather crowded. Although we are able to narrow it if we want to, and this ability is used in some languages, its role in speech is generally small. Above the pharynx, the vocal tract divides. One passageway goes up into the nasal cavity from which air escapes through the nose; however, this only happens if we allow it. We can close off the access to the nasal cavity by raising the **soft palate** (also known as the **velum**), or allow air to go into the nasal cavity by lowering it. The extreme end of the velum is a small piece of tissue called the **uvula** (you can see it by looking into the back of your mouth in a mirror—it is not a pretty sight), which plays a part in the pronunciation of some languages. Inside the mouth there are many parts which are used in speaking. The most obvious one is the tongue, which is so important in speaking that in many languages the word for 'language' is also the word for 'tongue'. It is amazingly mobile and flexible: consider, for example, how far you can push your tongue outside your mouth (some people can touch the tip of their nose with the tip of their tongue). This is done with muscles, yet muscles can only pull, not push. The tongue can move upwards and downwards, forwards and backwards. In producing vowels, the tongue does not make a great deal of contact with the **palate** (the upper surface of the mouth, sometimes called the 'roof' of the mouth); in many consonants, there is extensive contact. The lower jaw can also move upwards and downwards, and to a small extent forwards and backwards too. The teeth can be important in speaking, though we can't move them.

The outer end of the vocal tract is formed by the lips, which like the tongue, are very flexible and manoeuvrable. They can be moved towards each together and firmly closed, or can be moved further apart. They can be pushed forwards and rounded, or pulled back and widened as in a smile.

Describing speech production

Now that the various parts of the vocal tract have been introduced, let us look in detail at an example of how a word is produced. We will take the word 'sand'. You should read the description and see if it agrees with how you feel you say it. You will certainly find it strange to think of such a simple bit of speech requiring so many actions and such careful coordination. In speaking normally, we never have to think about what we are doing. It is very different for people recovering from a stroke, when with the help of a speech therapist they have to re-learn how to make even simple speech sounds.

You would probably take a breath before starting to speak, since there has to be enough air in your lungs to produce speech. The velum is then raised so that air cannot escape through the nose. The first sound, /s/, is made with the vocal folds apart and not vibrating. The air passes up through the larynx and the pharynx and into the mouth. Here it meets an obstacle: the front part of the tongue has been raised so that it is touching the roof of the mouth just behind the upper front teeth. This obstruction does not block the escape of air completely, but it forces the air to pass through a narrow gap and hit the teeth, making a hissing noise. The next sound is /æ/, and for this your vocal folds must be vibrating. So to move from /s/ to /æ/, you must bring the vocal folds together until they are in contact with each other, or nearly so. At the same time the tongue must be lowered so that it is no longer obstructing the flow of air. To produce an /æ/ sound the tongue must be quite low in the mouth, and the jaw is usually lowered. The velum is still raised to prevent the escape of air, but as you approach the /n/ sound which follows, the velum must be lowered, and this begins to happen towards the end of the /æ/. To make the /n/, the tongue and jaw must be raised into almost the same position it was in for making the initial /s/ sound, but this time the closure between the tongue and the roof of the mouth is complete. The vocal folds continue to vibrate, but now, with the velum lowered and the escape through the mouth blocked, all the flow of air passes through the nasal cavity and escapes through the nostrils. The final sound is /d/. The move from /n/ to /d/ is a very simple one: remember that during the /n/, the air is escaping

through the nasal cavity and no escape of air is possible through the mouth. You now raise your velum (most people are not aware of doing this when they speak), and the escape of air through the nasal cavity is stopped. The flow of air stops and the vocal folds soon stop vibrating. This raising of the velum results in the /d/ sound. Once this has been done, it is necessary to return to the position for normal breathing (you would not live very long if you stayed making the /d/ sound and did not start the flow of air again). The vocal folds are moved apart, and both the tongue and the velum are lowered, so that air can flow freely into and out of the lungs.

This description sounds very complicated; it may help in understanding the process to read through the diagram in Figure 2.2. We have looked at just four speech sounds: /s/, /n/, and /d/ are consonant sounds, while /æ/ is a vowel. Although most people have heard of vowels and consonants, they usually find it difficult to say what the difference is. In phonetics, the essential difference is that consonants make some obstruction to the flow of air, while vowels make relatively little obstruction. This means that if you produce alternate consonants and vowels (as in the word 'potato'), your vocal tract is changing from closed to open, open to closed alternately. However, although the consonants in 'potato' are clearly different from vowels, there are quite a lot of cases where the difference is not so easy to see, and the readings for this chapter reflect some of the different ways of dealing with this problem.

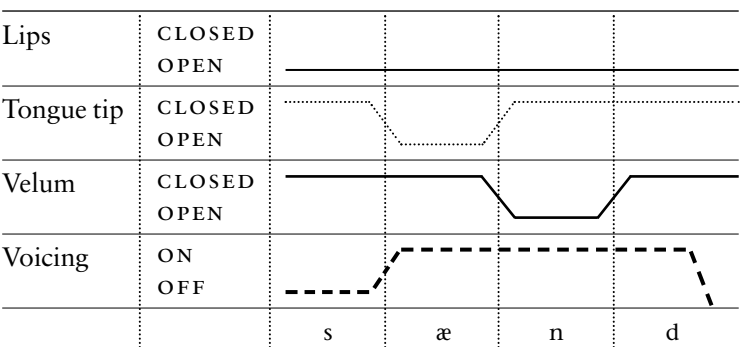

FIGURE 2.2 *Diagram of articulator movements for the word 'sand'*