THE PHYSIOLOGY OF SPEECH PRODUCTION

THE RESPIRATORY MECHANISM

As we have seen, sound consists of variations in pressure caused by the movement of molecules. Naturally, in order for these to start moving, there must be a source of energy, or **initiator**. In human beings, this is made up of the **respiratory** (i.e. breathing) **organs**.

In most cases, speech sounds rely on air supplied by the **lungs**, which are two spongy cone-shaped bags composed of tissues within a cavity (known as the *chest cavity, thoracic cavity,* or *thoracic cage*) enclosed by the **rib cage** and the **diaphragm** – a dome-shaped muscle that supports the base of the lungs and separates the **abdomen** (the belly) from the **thorax** (chest). Each lung is connected to the **trachea** (see below) by means of a small tube (*bronchus*). The left and right *bronchi*, in turn, branch off into a highly dense network of ever smaller tubes which ultimately end in *air sacs* (*alveoli*), where the oxygen is extracted and carbon dioxide released. In total, both lungs contain a total of some 750 million *alveoli.* Together, they represent some 70 square metres of surface – i.e. roughly the size of a tennis court.

During **inhalation** (or *inspiration*), the **diaphragm** contracts (and moves down), whereas the rib cage is elevated as a result of the action of the muscles between the ribs (the so-called *intercostal muscles*), thus enlarging the **thorax**. As the compartment enlarges, the lungs expand, which, in turn, results in the pressure inside the lungs dropping below the air pressure that surrounds us (known as the *ambient* or *atmospheric* pressure). It is through the inflow of air that the two pressures are equalized.

Conversely, when we breathe out, i.e. during **exhalation** (or *expiration*), there is a downward movement of the rib cage and a relaxation of the **diaphragm**, which then returns to its normal dome shape; this results in a reduction of the thoracic cavity, and the *compression* of the lungs, which, in turn, leads to an increase in the pressure inside them. And as the pressure inside the lungs is higher than the atmospheric pressure, air flows out in order to equalize both pressures.

Practice :

You can feel this process by sitting upright and placing the palm of one hand firmly on the stomach. Then, breathe in deeply, and breathe out again. During *inspiration*, you feel that the stomach expands, while it flattens again during *expiration*. You can also feel the expansion of the rib cage by repeating the previous exercise. This time, you should put the palm of one hand firmly on your chest ; when you breathe in, there is an outward and upward movement of the chest, with the opposite taking place when you breathe out again.

The movement of the **diaphragm** when you breathe in can also be observed very effectively during *hiccups*, which involves a spasmodic contraction of the **diaphragm**.

The entire process of breathing in and out is known as the **respiratory cycle**. At rest, this takes place some 12-15 times a minute, while it increases to thirty times during strenuous exercise, etc. In both cases, *inhalation* and *exhalation* tend to last equally long, and be equally deep. However, during **phonic respiration**, i.e. our breathing during speech, the *inhalation* time is reduced (sometimes to as little as half a second), whereas the *exhalation* time is increased to about 5-10 seconds, though in rapid, excited speech – when we wish to convey more speech – it can go up considerably. Trained singers, for instance, can make it last for up to 45 seconds – in exceptional cases even up to a minute! This corresponds to an average of 250-300 syllables in normal speech, and up to 400 in times of great emotional excitement or distress.

If you were wondering why **phonic respiration** relies on **exhalation**, try the following little experiment: say the sound '*aaah* (the sound a doctor asks you to make in order to examine the throat) and keep it going for as long as you can without breathing in (as the seconds wear on you will of course observe changes in the quality of the sound as more effort is required to 'squeeze' the air out of the lungs to maintain it). Next, breathe out entirely and say the sound '*aaah'* while *inhaling*. Again, keep it going for as long as you can. You will notice that this '*aaah*' cannot be sustained nearly as long as the *exhaled* one, whereas there is also a clear difference in the quality of the sound. As you have just experienced for yourself, the former method is a much more effective way of producing speech! For instance, if you try to utter an entire sentence – even a small one like 'Ann was here'– on *inhalation*, you will find that you do not get very far, while the result sounds distinctly eerie!

THE LARYNX AND VOCAL FOLDS

Before we can start producing speech sounds, the air passes through several stages of the respiratory tract, such as the **trachea** and the **larynx**, after which it goes through the so-called **vocal tract** (see below).

The **larynx** (< Greek λ ρυγξ, ‘throat’) is also known as the *phonatory system* and is located at the upper end of the **trachea** (or *windpipe*). It is a tube of 10-12 cm. long and about two centimetres in diameter, which consists of 16-20 horseshoe-shaped cartilage rings that are open at the back.

The **larynx** is popularly known as the **Adam’s Apple**. It lies immediately below the root of the tongue and is composed of a framework of cartilages, connected by means of joints and fibrous bands (*ligaments*) and membranes. The **larynx** has both *speech* and *physical* (i.e. anatomical) functions. Its main physical functions are to protect the entry to the lower air passages, and to prevent food, etc. from going down the **trachea**. Indeed, if any foreign substances reach the larynx, the **cough reflex** is the body’s way of trying to push them back up. Another *physical* function of the larynx is that it closes off the **trachea** (thus blocking the air inside the lungs) when the latter is stiffened together with the chest muscles during strenuous physical activities like lifting heavy objects, or childbirth. Finally, the larynx also plays a role in swallowing as it moves upwards and forwards under the base of the tongue, while at the same time the muscles of the **pharynx** (see below) push food, etc. towards the **oesophagus** (or **gullet**) – a tube of abot 25 cm in length which lies behind the trachea and runs parallel to it – so that it can continue towards the stomach.

Naturally, what concerns us here is the **speech function** of the larynx as it acts as the sound producing organ of the vocal apparatus. For this purpose, we need to discuss its three major components.

The larynx includes two big cartilages in the larynx. The lower one is called the **cricoid** cartilage, which is situated just above the highest ring in the trachea. It has a typical signet-ring shape (which explains its origins from the Greek for ‘ring’) – a narrow band at the front and sides, and a taller, wider part at the back.

Resting on this cartilage is the so-called **thyroid** cartilage (< Greek word for ‘shield’), which is made up of two square-shaped cartilages that are joined at the front of the larynx. At the top of where the two surfaces join together, there is a small V-shaped gap.

This is actually what is known colloquially as the *Adam’s Apple*. The outline of the front of the **thyroid** cartilage can easily be felt by tilting your head backwards and gently running your finger along the top of the Adam’s Apple. At the top and bottom, each wall of the **thyroid** cartilage extends into horn-shaped protrusions – the so-called *cornua*. The bottom ones fit over the **cricoid**, whereas the ones going upward are attached to the **hyoid** bone, to which the root of the tongue is anchored. When we swallow, the **hyoid** bone, the tongue, and larynx all move upwards. If you move your finger up the Adam’s Apple past the V-shaped notch at the top of the **thyroid** cartilage, you will at some point feel a bony ridge. This is the front part of the **hyoid** bone.

Inside the ‘box’ formed by the **thyroid** and **cricoid** cartilages we find the **vocal folds** (**- lips/ - cords**), i.e. two three-sided pyramid-shaped muscles, which project into the cavity of the larynx. The older term *vocal c(h)ords* is a literal translation of the French *corde vocale*, introduced by the French physician Ferrein (1741), who was the first to perform laryngeal experiments. His terminology was linked to his belief that these organs acted like the strings on a musical instrument. As we shall see, however, this theory does not reflect reality.

The **vocal folds** are stretched across the larynx from front to back. The opening between the vocal folds is called the **glottis** (< Gr. ‘tongue’). The inner edge of the vocal fold has a length of 17-23 mm in men, and about 13-18 mm in women. At the front they are joined together and fixed to the inside of the **thyroid** cartilage, whereas at the back they are attached to a pair of small cartilages, the **arytenoid** cartilages, whose name is derived from the Greek word for ‘ladle - in reference to their shape. These are arguably the most vital organs of speech since it is they that move the **vocal folds**; the **arytenoids** articulate with the **cricoid** through a joint which allows them to rotate and slide, and thus to open and close the **glottis**.

Let us now take a look at the sounds that correspond to the various positions of the vocal folds. When you are breathing normally, or producing a sound like **/f/**, they are wide apart (**abducted**). As we have seen, they can also be tightly closed (**adducted**), for instance, when we are lifting heavy objects. In speech, the **glottis** is sometimes completely closed and then opened suddenly to release air that has been compressed by pressure from the lungs; the resultant sound is known as a **glottal stop** (**/ʔ/**), which will be discussed later on.

Far more spectacular, however, is the process involving the **vibration** of the vocal folds. It is possible to distinguish four stages in the mechanism. The first is **closure** (or **adduction**), during which the laryngeal muscles bring the vocal folds together. As the flow of air from the lungs continues, there is a build-up of air below the folds (= **subglottal pressure**). This stage is known as **compression**. At some point, the compressed air will force apart the vocal folds so that a little air escapes for a brief moment. This phase is the **release**. As the air flows past, the vocal folds are brought back together as a result of two forces: the elasticity of the folds, and the so-called **Bernouilli effect**. The air rushes through the narrow opening in the vocal folds at great speed (faster than the usual outflow speed from the lungs). As a result, the pressure between the vocal folds drops (in relation to the pressure above and below them). It is the force of this negative pressure which sucks the vocal folds back together. This phenomenon has been named after the Swiss physicist Daniel Bernouilli (1700-1782), who discovered a similar process during his experiments with fluids and gases. The **Bernouilli effect** can also be observed in a number of other cases. The easiest way is to take two sheets of paper and to hold them so that they hang vertically two centimetres or so apart. Then, blow downwards so as to produce a current of air, and you will see that they are drawn together as a result of the reduction in pressure associated with the current. The same phenomenon is also responsible for the drawing together of two ships moving through the water in the same direction at the same speed with a small distance between them. In this case, the current results from the displacement of water by each ship's bow.

When the vocal folds are brought back together again, the subglottal pressure again builds up, and the process repeats itself. The entire process of opening and closing of the vocal folds is called the **glottal cycle** by phoneticians. On average, this occurs between 200 and 300 times per second (expressed as *cycles per second*) in women, and about 100 to 150 times per second in men. It is important to add that the vibration rate does not remain constant in speech, and is controlled by the speaker who changes it in order to achieve certain perceptual effects. An increase in the frequency of this process leads to an increase in the **pitch** of the voice, and vice versa. You can experience this for yourself by starting out with a very low ‘aaaaa’ sound and then gradually raising the pitch; what you are doing is simply increasing the vibration rate of the vocal folds.

This rapid vibration of the vocal folds results in a typical **buzzing** sound, which we call **voice**, or **phonation**. This can be heard in consonants like **/z/** or **/b/**, which are then said to be **voiced**. In English, for instance, all **vowels** and **nasal consonants** (e.g. **/m/**) are voiced. When the folds are apart – leaving space for the air to pass without obstruction – specialists use the term **nil phonation**, and any speech sounds produced without the vibration of the vocal folds are **voiceless** (or *aphonic*): e.g. **/p/, /t/**. The presence or absence of voice can best be detected by sticking your index fingers in your ears; voiced sounds then produce a loud buzzing noise. Another way of detecting voice is by holding your index finger and thumb to your **larynx**, which allows you to actually feel the vibration when you produce a voiced sound.

Practice :

1. Cover your ears and say the sound /ffffffff/. Then, say the sound /vvvvvvvvv/. The buzzing you hear in the second sequence is the vibration of the vocal folds. Now alternate between /fffffff/ and /vvvvv/ ; what you are doing now is switching your vocal folds ‘on’ and ‘off’.
2. Repeat the same exercise, but this time place your thumb and index finger at the sides of your Adam’s Apple. This allows you to *feel* the vibration of the folds.

The **pulses** corresponding to vibrations of the vocal folds in speech sounds can be visualized through a so-called **wave form**, which is a picture of the variations in air pressure associated with these sounds.

In addition to the voicing/voicelessness described here, there are other types of phonation, which will be discussed later on. At this stage, however, it is important to know that the **vocal folds** can also be subject to other changes. For instance, the extent to which they are pressed together, or whether they are tensed or relaxed all play a major role in the quality of the resultant speech sound, as do the pressure of air below the folds, the force with which the air is pushed through (**intensity**), the **frequency** at which they vibrate, and their **length**. For instance, the generally higher **pitch** of women’s voices in comparison to men is the result of the smaller size of their vocal folds and a higher average rate of vibration.

Above the vocal folds, there are two other pieces of tissue, which are known as the **false vocal folds**, or **ventricular bands** (or **folds**). These are not usually used in normal speech, except for special effect. Their use generally means that the speaker has a voice disorder. Probably the most famous example of a voice quality of this type is that of the American jazz musician Louis Armstrong.

As mentioned earlier, there are a number of types of phonation. Up until now, we have only talked about **true** (or **modal**) **phonation**, i.e. that used in the production of **voiced** speech sounds in English. However, the **vocal folds** can operate in a number of different ways, resulting in different types of **phonation**. A few of them will be discussed here.

1. **Murmur** (or **breathy voice**)

This type of voicing involves the vibration of the vocal folds while they are **apart**. It is characterized by a large escape of air during vibration, which gives rise to audible friction. This is used in some languages spoken on the Indian Subcontinent (e.g. Hindi, Gujarati), but can also be heard in the intervocalic (= between vowels) **/h/** in English words like *ahead*. Although the IPA denotes this type of voicing by two dots underneath the symbol (e.g. [**[a̤]**), the sound in *ahead* is usually transcribed by the symbol **/ɦ/** (e.g. **/əˈɦed/**), in narrow transcription.

In English, and many other languages, **breathy voice** is often used paralinguistically in intimate contexts. It can also be heard, for instance, in television adverts to underline the softness or smoothness of things like sheets or toilet paper. Some people always have some degree of breathy voicing in their speech. In women, in particular, this kind of 'husky' voice quality is generally thought to be sexy.

2. **Creaky voice** (**vocal fry, laryngealization**)

In this case, the **arytenoid** cartilages are held closely together, so that only the **anterior** (forward) part of the vocal folds are vibrating. This type of voicing (which is marked by a very low pitch) is used phonetically in some American Indian languages, as well as some in Africa (e.g. Hausa) and Asia (e.g. Vietnamese), while in English it can be heard at the end of falling intonations of speakers with a very traditional RP speech. This type of phonation has also been compared to 'a series of rapid taps, like a stick bein run along a railing' (Catford 1962). The IPA diacritic for **creaky voice** is a tilde-shaped line underneath the symbol: e.g. **[b̰]**.

Practice :
Pronounce an /aaaaaaaaaaaaaa/ sound and lower the pitch until you cannot get any lower. If you then continue, you hear the regular taps of the vocal fold vibration.

3. Whisper

This type of phonation involves keeping the vocal folds rigid and close together, except for a triangular opening caused by the posterior (i.e. back) edges of the **arytenoid** cartilages being drawn apart. The pulmonic air is forced through this narrow slit, which results in the typical ‘hissing’ sound associated with a **whisper**. It should be noted here that whisper affects all sounds in a given utterance. The IPA **diacritic** for whisper is a full stop underneath the symbol: e.g. **[a]**.