We saw in chapter 3 on respiration that many anatomical structures and functions necessary to speech production have more basic life-sustaining functions. This is also true for the vocal folds. Phonation refers to the vibration of the vocal folds.

The Vocal Folds

Although the primary function of the vocal folds is to act as a valve for the lungs and protect them from the aspiration of food and other substances into the lungs, they are also essential for the productions of speech. Sometimes they are referred to as the "vocal cords." However, they are attached along one side and do not vibrate freely like a cord, so we will use the preferred term vocal folds.

The vocal folds can be approximated tightly to keep air inside of the lungs and make the thorax (chest cavity) rigid and provide a basis of support for lifting heavy objects. (You might have guessed that the vocal folds were somehow involved in lifting heavy weights from the grunts of weightlifters or the sound that some tennis players make when their racquets make contact with the ball.)

Everyone is familiar with the role that the vocal folds play in expelling substances out of the lungs by coughing. Coughing is a rapid release of subglottal air pressure that can be made intentionally or unintentionally. The purpose of a cough is to blow substances such as mucus away from the entrance to the lungs. We take such a basic act as coughing for granted. However, persons with dysphagia (a disorder of swallowing) may not be able to cough properly, making them much more susceptible to choking and pneumonia due to aspiration of food or liquid into the lungs. Indeed, our entire species is vulnerable to choking because the openings to the lungs and to the stomach are relatively close together. "Talking with your mouth full" may result in an opened trachea and esophagus, and many people have died from food passing down the wrong tube.

In contrast, no other species has such an accident-prone vocal anatomy, and even human infants start with a safer vocal design than adult humans. The payoff for having our vocal folds close to the opening to the esophagus, though, is that our entire vocal tract is larger due to a lower larynx. Because of the nearly right-angle bend in the modern human's vocal tract, we can better isolate two independently
resonating cavities (one oral and the other pharyngeal) and thus produce more and better speech sounds. Some of the speech sounds that are allowed by the human vocal tract are particularly acoustically stable sounds. These sounds are referred to as voiced. Voiced sounds allow for faster speech production because they can be articulated with relative insensitivity with little change in their acoustic structure. When the vocal folds are placed under just the right conditions—they are approximated firmly but not too tightly—they can vibrate rapidly. It is this rapid vibration of the vocal folds that we hear as "voice." You can feel this vibration by humming and placing your fingers on your larynx (or "voice box"). Because males have somewhat larger larynges, this area of the throat is a little more prominent on males, where it is called the "Adam's apple."

There are actually two senses of the term voice. One is nontechnical and refers to the individual sound quality of a particular person when he or she speaks. The other more technical sense of the term refers to the sound that the vocal folds make when they vibrate. This latter sense of the term refers to the vibrating sound source for many speech sounds.

**Fundamental Frequency**

The rate at which the vocal folds vibrate is referred to as the fundamental frequency. Fundamental frequency is usually perceived as a person's vocal pitch, so a lower \( f_0 \) value is heard as a deeper voice. Most women and all children have higher \( f_0 \) values, which listeners hear as higher-pitched voices. It should be pointed out, however, that the vocal folds are constantly changing in fundamental frequency from when they first begin to vibrate, as the speaker places more emphasis on certain words and as the vocal folds slow down for the end of a sentence.

**Hint to Students**

Fundamental frequency refers to the actual rate of vibrations of the vocal folds, while pitch refers to the perception of fundamental frequency by human listeners.

**Hint to Students**

Originally the symbol for the fundamental frequency was \( f_0 \) (the integral sign, sub-zero). However, probably because of typesetting limitations, the capital \( F \) symbol with a zero \( (F_0) \), or \( F_0 \), has become the standard in most texts. This is somewhat unfortunate, because it makes it much easier for students to confuse the fundamental frequency with formant frequencies which are written \( F_1, F_2, F_3, \) etc. Since, as we will see, there is no relationship between the fundamental frequency and formant frequencies, even though both are frequency measures expressed in Hertz, we will return to the use of \( f_0 \) in this text. We will do this even though students should be aware that many other texts use \( F_0 \).
Phonation

So even though a single average value is typically used to describe the fundamental frequency, you should be aware that the rate of vibration of the vocal folds is not static but rather is constantly changing. Typically this rate of vibration is measured in Hertz. Hertz, abbreviated Hz, has replaced the older term cycles per second or c.p.s. There is one Hertz for each cycle of each complete opening and closing of the vocal folds per second. So for an adult male voice with a fundamental frequency of 100 Hz, the vocal folds open and close completely 100 times in one second (although this value is a little low for a typical male voice; perhaps 120 Hertz is a more typical value). An average adult female might have a fundamental frequency of 200 Hz, so her vocal folds would open and close completely 200 times per second.

**Hint to Students**

Like the speed of driving a car in miles per hour, the vocal folds vibrate at different rates in Hertz. Even though the average speed of an automobile may be 40 miles per hour, we constantly have to slow down and speed up again at traffic lights. Similarly, there are voiceless sounds in most sentences that require the vocal folds to momentarily stop vibrating.

For a child, the vocal folds may vibrate at 300 times per second (i.e., 300 Hz) or higher. Although there is no single agreed upon average values for each gender, these are somewhat convenient values—100 Hz for men, one octave (an octave is a doubling of frequency) higher at 200 Hz for women, and another 100 Hz higher for infants at 300 Hz.

**Hint to Students**

Larger structures vibrate more slowly than smaller structures. Think of larger, thicker vocal folds taking longer to complete one cycle of phonation.

It should be pointed out here that even though there are some differences in the typical fundamental frequency (rate of vibration of the vocal folds) for different differences in the vocal folds and laryngeal structure due to age and gender, it is not these differences that account for the differences between the various sounds of speech. The fundamental frequency can remain the same for a whole series of vowel sounds produced by the same speaker. The simple sound of the vibrating vocal folds is modified through resonance and molded into the various sounds of speech. This process of resonance is considered in the next chapter.

In order for the vocal folds to vibrate, the speaker must have inspired or taken sufficient air into the lungs. Remember that one takes in more air during an inspiration that is going to be used for speech than would be the case in quiet breathing.
When breathing quietly, a person typically spends about equal time inspir- ing (breathing in) and expiring (breathing out). But a person who is speaking typically spends as little as 10 percent of the breathing cycle in inspiration and as much as 90 percent or more breathing out. This is understandable since, in almost all cases, speech is produced during expiration—what is called an egressive flow of air. Although it is possible to produce some speech on inspiration (the inward or ingressive flow of air), this is usually very rare. In some languages, speakers may communicate the last unit of a series of items on inward flow of air in order to emphatically signal the end. Or they may do so in order to emphasize certain words.

It is interesting to observe just how skilled speakers are at coordinating respiration and speech, since they rarely (almost never) run out of breath before the end of a sentence. Lieberman and Lieberman (1972) found that people have an idea of the length of their intended utterance and inspire accordingly. Speakers typically only take breaths in between sentences, and they may even produce two or three sentences on the same breath. There are various forms of evidence that have been used to demonstrate the rather intricate control of breathing when its purpose is speech. (Recall from the previous chapter that speakers tend to produce questions that require a rise in fundamental frequency at a faster rate than statements that do not require the rise.)

The Anatomy of Phonation

Before detailing the process of phonation, we must briefly review the anatomy of the larynx, which is the series of cartilages, muscles, and tendons that make up and surround the vocal folds. The vocal folds are associated with three main cartilaginous structures and muscles that either close the vocal folds, open them, or tighten them. A speaker may increase tension on the vocal folds in order to raise the rate of vocal fold vibration, and hence raise the f0.

As can be seen in Figure 4.1, the three main cartilages of the larynx are the thyroid cartilage, the cricoid cartilage, and a pair called the arytenoid cartilage.

We will discuss each in turn. The thyroid cartilage looks like a shield with the fuller section in front and thinned section in back. At the front, this shield arches, and for men this arch is at a more acute angle than for women. The male Adam's apple, prominent on men's throats, results from the notch of the thyroid cartilage. Women have Adam's apples in a way, in that there is still a thyroid notch, but it is not as acute as in men, and hence less pronounced.

Hint to Students

The thyroid cartilage is like a shield; the cricoid like a ring; and the arytenoids are like triangles.
Above the thyroid is the cricoid cartilage, which is shaped like a signet, or school ring. Unlike a ring, though, the fullest section is found in back and the thinner part anterior to (in front of) the larynx.

Sitting on top of the cricoid are the two triangular shaped arytenoid cartilages. They can rotate inward and outward when certain muscles are tautened. Generally, all parts of the larynx and the structure move as a whole, mainly up and down. The structure known as the vocal folds stretches from the arytenoids to the inside of the thyroid cartilage. Because of this, the movement of the whole larynx may result in a higher fundamental frequency. Specifically, the vocal folds attach to an angle of the arytenoid called the vocal process. The folds themselves are composed of muscles, tendons, and ligaments.

There are five main muscles that alter the placement of the vocal folds. The names of these muscles will help you remember where they are located, and vice versa. The three cartilages and these muscles all work together to prepare the vocal folds to vibrate, remain still, or tense for the purposes of increasing voice pitch. These muscles are illustrated in Figure 4.2. Although there is not complete
agreement in the scientific literature on exactly which muscles are responsible for which laryngeal adjustments, the summary that follows is consistent with most accounts.

The Process of Phonation

Closing the vocal folds is called adduction. The interarytenoid muscle runs between the two pools of the arytenoid cartilage. When muscles are tensed, they shorten. Tensing the interarytenoid muscle will pull the two triangles of the arytenoid cartilage together. This maneuver also brings the vocal folds closer together.

A secondary adducting muscle is the lateral cricoarytenoid muscle. Some speech sounds—for example, stop consonants—require a great deal of intralaryngeal pressure. That is, they need air pressure in the oral cavity (mouth) to be great, with no leakage. An extra tight vocal fold adduction helps seal off the pharyngolaryngeal cavity. In such a case, both adduction muscles would be employed. The lateral
cricothyroid muscle is located on the sides (lateral) of the cricoid and stretches to the arytenoids. When this muscle tenses, the shortened lateral cricoarytenoid pulls a corner of the arytenoid triangle (the muscular processes) forward, thus rotating both cartilages of the arytenoid. This rotation movement pulls the vocal processes closer, since the vocal folds are attached, they too approximate.

Pulling the vocal folds apart is called abduction. The posterior cricoarytenoid muscle is attached in the back (posterior) between the cricoid cartilage and the muscular processes of the arytenoid cartilages. When tensed, this muscle rotates the muscular processes inward and closer together. The vocal processes thus separate; consequently, the vocal folds also separate.

Tensing the vocal folds involves a primary and a secondary muscle. When a mass is pulled taut, it is thinner at any specific point and will thus vibrate more quickly. This is one way to increase the fundamental frequency (heard as pitch) of one’s voice. To raise the fundamental frequency, the cricothyroid muscle can be employed. This structure connects between the cricoid and thyroid cartilages and when used will actually rock the cricoid back, pulling the arytenoids with it. The vocal folds tense up as a consequence, affecting the fundamental frequency, which increases.

The thyroarytenoid muscle is a secondary tensing muscle. This muscle, also called the vocalis, runs the length of the vocal folds themselves, between the thyroid and arytenoid cartilages. When tensed, this muscle works to decrease the vibrating mass of the vocal folds. Consequently, vibration actually increases. Remember, a thinner object can move more quickly, and the tensed vocal folds are technically thinner, although they also increase slightly in length.

**Hint to Students**

Primary and secondary muscles altering vocal fold placement and tension can be remembered as follows: abduction by posterior cricoarytenoid (PCA); adduction by infraarytenoid (IA), with additional closure by lateral cricoarytenoid (LCA); tensing by cricothyroid (CT), with additional tensing by thyroarytenoid (TA).

The fundamental frequency can actually be affected in four ways by various processes. We have already discussed two ways to increase tension along the vocal folds in order to increase $f_v$. Primarily, we can employ the cricothyroid muscle, and secondarily the thyroarytenoid muscle. We can also increase fundamental frequency by increasing subglottal air pressure. When a greater discrepancy exists between air pressure below the vocal folds and above the vocal folds, they will be blown farther apart. This will increase intensity or loudness. So amplitude and $f_v$ rise together when subglottal air pressure is increased.
However, these two acoustic parameters can be manipulated separately. As a matter of fact, studies have shown that people stress syllables with individually varying combinations of amplitude, $f_o$, and duration manipulation, sometimes using all three or only one correlate to signal stress in a word (Deherra, 1988; Lieberman, 1960).

**Hint to Students**

Speakers can alter amplitude and fundamental frequency independently, so that an increase in one can, but need not, be accompanied by an increase in the other.

Finally, speakers can lower their $f_o$ by lowering the entire laryngeal structure. This movement adds mass to the vocal folds and can be executed with the help of strap muscles in the neck. The vocal folds are apart to let in the free flow of oxygen during inspiration. They must be drawn together or approximated (also called adduction) in order for them to begin vibrating.

**Hint to Students**

To remember adduction versus abduction, think of adding as a process of putting things together. In this case, the vocal folds while a person who is abducted is carried away.

Because the vocal folds are closed, air pressure builds up under the vocal folds. When this subglottal air pressure reaches a certain critical level, so that it is greater than above the glottis (i.e., supraglottal), the vocal folds are pushed apart and the air pressure from beneath then escapes.

Two forces act together to bring the vocal folds back together again for the cycle of vocal fold vibration to begin once more. First of all, the vocal folds are made up of resilient muscle and ligament material—that is, they are inherently elastic. Something else that is stretched has a tendency to return to its natural unstretched position. But there is also an aerodynamic principle, called the Bernoulli force, that acts to draw the vocal folds back together again. Simply put, the Bernoulli force is created from the release of air pressure when the vocal folds are first blown apart. The air rushes out faster, and this drop in pressure between the vocal folds causes the suction-like Bernoulli force, resulting in vocal fold adduction. In an airplane it is actually the Bernoulli force that causes the plane to lift off the ground. The Bernoulli force results from air pressure being greater under the
wing of an airplane than over the wing since the air underneath is rushing by faster. The design of the wing makes it such that the air has a longer distance to travel under the wing than over the wing. This, in turn, causes wings to lift a plane off the ground. 

You may also experience the effects of the Bernoulli force when driving close to the rear of a fast-moving truck and feeling your ear being pulled toward it. But for speech purposes, you simply need to understand that the drop in air pressure caused by the opening of the vocal folds also creates the Bernoulli suction force that helps to bring them back together again.

These two forces that serve to bring the vocal folds back together again—one muscular—elastic, the other aerodynamic in nature—explain why this account of vocal fold activity is termed the myoelastic aerodynamic theory of vocal fold vibration.

**Hint to Students**

myoe= muscle, elastic= elastic recoil, aer= air, dynamic= movement.

A somewhat more detailed account of vocal fold vibration takes into account that the vocal folds are not one undifferentiated mass; rather, there are layers to the vocal folds. The vibration of the vocal folds is more accurately represented by a model in which there is a stiffer body that is covered by a more flexible layer. Hence this model is known as the cover-body theory of vocal fold vibration (Bhatnager & Kakita, 1985).

There are also many fine details of vocal fold vibration, such as whether the vocal folds are entirely closed along their edge or whether a small gap or "chink" remains, giving a person's voice its individual character. It is rather amazing to think that there are enough individual differences in vocal fold vibration (along with other speech processes) to give voices their individual identity. We are usually able to recognize a voice over the telephone, even one that we haven't heard in years. And, of course, we can recognize the voices of all sorts of people we interact with on an infrequent basis. We often take this ability in listeners for granted and do not always bother to state our own name to a close friend or relative over the phone: "Hi, it's me."

It only makes sense that individual voice recognition had a survival value in early humans, which may explain just how developed voice recognition is in our species. There is also evidence that this skill emerges very early in life, and many would argue that at least a part of voice recognition is innate. It's otherwise difficult to explain why a newborn infant shows a preference for its parents' voices from the time of birth. Certainly some of the individual acoustic character of a voice can be carried through the amniotic fluid of the womb to the ears of the developing fetus, but the fact that the future child knows what to listen for suggests
injate genetic support for this process. More about the emergence of language abilities in infants can be found in chapter 15.

SLPs have to understand phonation in order to serve their clients with various voice disorders. Problems in voice can range from the simple temporary hoarse voice resulting from a cold or vocal abuse (such as extensive shouting at a sporting event or cheerleaders who tend to shout while performing strenuous physical exertation, perhaps one of the quickest routes to vocal problems) to more complex problems such as vocal nodules (which prevent the vocal folds from closing completely), to laryngeal cancer and even removal of the laryngeal mechanisms (known as a laryngectomy). SLPs also treat secondary problems of laryngeal valving such as can occur in swallowing disorders. Some of the methods for observing the activity of the vocal folds is considered in chapter 10 on physiological measures of speech.

To the next two chapters, we will discuss how air from the lungs, after encountering the vocal folds, passes through the array of human articulators and cavities of the vocal tract to resonate and result in speech.

Study Questions

1. How are the vocal folds adjusted to produce voiced sounds? To produce voiceless sounds?

2. How may speakers increase their fundamental frequency? How may they increase voice amplitude? Explain the ways these two acoustic correlates are related.

References


