5. Sourcing Speech: Phonation

So far, we've learned about how a vibrating object interacts with air to give us a complex sound, and how this sound can be filtered by tubes with varying physical characteristics to amplify certain harmonic frequencies. We've also learned about the respiratory system and laryngeal structures in the throat that generate and support the production of complex sound. In combining these concepts we'll now discover how speech is *sourced* by our vibrating vocal folds. This action produces a complex sound that's ultimately filtered by the tubes in our head (the pharyngeal, oral, and nasal cavities).

In this chapter we'll direct our focus to this conversion from turbulent air rising up from the lungs to the structured acoustic energy generated by its interaction with the vocal folds. The sound that originates from the glottis is broadly called **phonation**, but we'll use it to refer to a particular type of sound—the complex periodic source sound resulting from the opening and closing of our vocal folds.[1] Phonation, more often called **voicing** in the linguistics literature, produces the loudest aspects of speech. All vowel sounds (aaaa-s, oooo-s, uuuu-s, etc), nasal sounds (m's and n's), and the voiced series of consonants (b, d, g, v, z, for example, but *not* p, t, k, f, s, h) are produced with the vocal folds vibrating. If we think about speech like the mechanics of a gas-powered car, the process of phonation would be comparable to internal combustion, with the vocal folds being the pistons that are set into motion. Let's look at how and why the vocal folds begin moving, and what effect their movement has on the air right above the glottis, before it is filtered and shaped by your pharynx, larynx, and mouth.

[1] In some literature *phonation* can refer to even the sound of turbulent air flow that doesn't result in vibrating vocal folds, such as in whispering or sounds like hhhh and ffff. But for most practical purposes defining *phonation* as vibrating vocal folds is reasonable.

4.1 Vocal fold vibration

The earliest (and simplest) modern **model** of how and why vocal folds begin (and continue) vibrating (you'll also see the term "**oscillating"**) is called the **myoelastic aerodynamic theory of phonation,** proposed by Janwillem van den Berg (who also patented the implantable pacemaker!) in the late 1950s. I call the myoelastic theory the "simplest" model because numerous subsequent models have expanded and refined the foundational proposal made by van den Berg. The name of the model tells you what is involved in phonation: *myo* ("muscle" in Greek), *elastic* (referring to the restoring force of the vocal folds), *aerodynamic* (relating to the pressure and flow of air). The interaction of muscles of the vocal folds with the pressurized air below the glottis results in phonation.

The basic process of phonation proceeds as follows:

1. The vocal folds must be in a closed (**adducted**) state,
2. The air pressure below the glottis (in the trachea and lungs), called **sub-glottal pressure** (P_{sub}), increases until,
3. The P_{sub} reaches a point (**phonation threshold** pressure) where the vocal folds can no longer remain adducted and begin to be forced apart.
4. Because of the pressure differential, the sub-glottal air, whose pressure is positive (greater than P_{am} above the glottis), escapes through the glottal opening and vented across the vocal folds.
5. The vocal folds, which have been pushed open begin to recoil, or snap back to their adducted position, because of the elastic properties of the vocal fold tissues.

6. As the vocal folds begin to adduct, they create a narrow channel, causing the air passing through to move faster and decrease in pressure below P_{am} (**Bernoulli Effect**).
7. The negative pressure air (P_{neg}) between the vocal folds further causes them to adduct, and
8. The process repeats.

In normal phonation, the opening **(abduction)** and closing **(adduction**) of the vocal folds occurs at a rate of around 200 times per second for females and around 100 times per second for males (though there is a *lot* of variability) (Zemlin, 1998). The oscillating vocal folds excite the air in the *supra-glottal* (above the glottis) cavity in much the same way that the tines of the tuning fork pushed surrounding air to create a tone (Chapter 1). While the tone from our tuning fork is simple, the sound resulting from the oscillating vocal folds is *complex* (Chapter 2). But before we describe all the complexities of the sound emanating from the **glottal source**, let's understand why the myoelastic aerodynamic process happens in the first place.

You can model steps (1)-(4) yourself by closing your lips tight and puffing your cheeks out. Now slowly puff them out more and more. At some point the air pressure inside your mouth will be difficult to contain and your lips will be pushed apart ever so slightly, making a *raspberry*-like sound, like the sound trumpet players make into the mouthpiece of their instrument. Step 5 may be a little difficult to visualize or analogize, so let's return to our balloon example. Imagine a fully inflated balloon with the neck pinched off with your fingers. If you were to keep the neck pinched and slowly pull apart the lips or bead of the neck, the high-pressure air inside the balloon would be released in much the same way as the positive P_{sub} air is pushes through the vocal folds. Pulling the lips of the balloon neck apart ever so slightly models

the subglottal air pressure overcoming or breaking through the adducted vocal folds. If we were to zoom in on the lips of the balloon and played it in slow motion, you would see them moving back and forth, like clapping hands. Step 5 says that when the air pushes the vocal folds apart, they bounce back due to their elasticity. In the balloon example the lips are likewise pushed apart when you let the air out. They return to their midline position (very briefly) because of the elasticity of the rubber.

Step 6 is a bit magical. The basic idea is that the vocal folds close not only because of their elastic property, but *also* because of the Bernoulli Effect, which says that pressure is inversely related to the velocity of air (or any gas or fluid). This is precisely the case with the air that is escaping the trachea through the closing vocal folds (which are recoiling due to elasticity). The closing vocal folds creates a narrow channel, which affects the velocity of the air travelling up from the trachea. The volume of air remains constant from the trachea and into the narrow channel between the vocal folds, as a result, the air moves faster or has a greater velocity than it did in the trachea.[2]

DRAWING OF WIDE TUBE CONNECTED TO NARROW TUBE WITH CROSS SECTIONAL AREA AND VELOCITY SHOWN

The Swiss physicist, Daniel Bernoulli, modeled what we now call the Bernoulli Effect or Principle, which states that an increase in fluid (or in this case air) speed occurs simultaneously with a decrease in pressure (velocity up, pressure down). The math and physics behind this effect

[2] This is called the *continuity equation*, which says the cross-sectional area (A_1) of the tube in which the fluid or gas is travelling, multiplied by its velocity (V_1), is equal to the cross-sectional area (A_2) of the narrower tube the fluid has moved in to, multiplied by its velocity (V_2) in that narrower tube. So, a decrease in A_2 results in an increase in V_2.

are complex, but the takeaway for speech is rather simple. With the mass of air from the trachea moving fast through the glottal opening or **aperture**, the pressure is decreased. The low pressure

in the glottal channel causes the sides or the vocal folds themselves to be forced further to midline or adducted position. When you let the air out of your bike tire, the pressure inside the tire decreases, causing the tire to collapse into itself. Similarly, the sides of the channel (or the vocal folds themselves) are forced to further close.

PICTURE OF CROSS SECTION OF VOCAL FOLDS WITH ARROWS REPRESENTING AIR FLOW AND DECREASED PRESSURE IN NARROW CHANNEL

Think about it!
Have you ever experienced the following situation: you're enjoying a shower when suddenly, the shower curtain creeps in towards you? Well, this occurs because of the Bernoulli effect—the fast-moving water from the shower head *entrains* (or brings along with it) the surrounding air. The fast-moving air (caused by the the fast-moving water) lowers the pressure in the shower relative to P_{am} outside the shower, causing the curtain to get sucked in—just like the vocal folds during phonation.

The myoelastic aerodynamic theory of phonation results in the vocal folds oscillating in a *self-sustaining* manner, that is, the vibration continues without any external forces being applied to the glottis. Interestingly, this is not the only proposed model of how phonation occurs. In 1950 the French *phoniatrist*, Raoul Husson, published the "neurochronaxic theory of vocal-cord vibration." Husson claimed that vocal folds don't vibrate

because of aerodynamics and tissue elasticity, but rather they vibrate as a result of neural stimulation and muscle contraction! The neurochronaxic hypothesis received lots of criticism and faded into the annals of misguided scientific theories.

4.2 The shape of vocal fold vibration

Now that you have a sense of *why* vocal folds vibrate, let's take a look at *how* they move. Although it's convenient to visualize the vocal folds as a set of swinging doors (like in a western saloon) being pushed open by sub-glottal air pressure, they have a much more complex shape and movement, perhaps more akin to a pair of flopping fish out of water. They don't open and close in one fell swoop, but rather progressively, with different parts of either side of the folds separating (during the opening movement) or making contact (during the closing movement) at different times. In the cross-section below, notice how the top portion of the folds initially separates when subglottal pressure has been achieved, then the lower portion closes because of the recoil and the Bernoulli effect. There is a small timing difference between the **superior** (upper) and **inferior** (lower) portions of the vocal folds. Likewise, there is a small difference in the timing of the opening of the folds along the **anterior**(front)/**posterior**(back) dimension, with the posterior portion opening before the anterior. The recoil sequence is just the opposite, with the anterior portion closing first.

ILLUSTRATION OF VARIOUS STAGES OF VOCAL FOLD OPENING AND CLOSING

The opening and closing of the vocal folds, and in particular the sequence of the movement of the various sections of the vocal folds, have inspired models of their dynamic

movement. These models consider the complex movement of the vocal folds as independently moving *masses* that work together to control the flow of air escaping the lungs. They also refine the aerodynamics of the original myoelastic theory. For example, in the **single-mass model** (Flanagan and Landgraf, 1968) each fold is modeled as a simple harmonic oscillator (moving in a uniform pattern like our tuning fork example in Chapters 1 and 2). The model also introduces the aerodynamic contribution of supraglottal pressures. When the mass of air travels away from the glottis (and into the pharynx) it leaves behind a region of low pressure in its wake. This low pressure further helps adduct the vocal folds as they recoil. In this way, researchers have modeled the vocal folds with more and more masses (some models have 16 masses!) accounting for ever greater complexity of the vocal fold movement. But for the purposes of understanding how we produce *voice,* the myoelastic theory gets us pretty far.

4.3 The sound of vocal fold vibration

The sound produced by the vibrating vocal folds is complex. There is the sound of the main vibration, which is the slowest and loudest, producing the lowest or fundamental frequency (F_0), and then there are the sounds of higher frequency harmonics, reflecting the faster and smaller vibrations along the length of the folds. Remember from Chapter 2, that the higher-level harmonics are related to the fundamental frequency in a systematic way—they are whole integer multiples of the fundamental frequency.

DRAWING OF VOCAL FOLDS SHOWING SMALL VIBRATIONS

Like other complex periodic sounds, the sound from the vibrating vocal folds is best visualized using the power spectrum (which when referring to vocal fold vibrations is called the **glottal spectrum**). The fundamental frequency is the most dramatic movement of the vocal folds, displacing the vocal folds to the greatest degree and therefore giving us the loudest harmonic, it's the frequency that tells the listener about the speakers pitch. Higher frequency harmonics (above the fundamental frequency, or to the left of the fundamental frequency in the spectrum picture) decrease in amplitude. What this means is that the displacement of the vocal folds gets smaller as the frequency increases.

The decrease in amplitude of the higher harmonics is captured by the term **spectral slope** (also called **roll-off)**, which is the rate at which the amplitude decreases. The slope is generally described by referring to the **octave**, which is a *doubling* (or *halving*) of any frequency. You may know the term from music, where a particular note can be played or sung at various octaves. For example, an A note, which has a fundamental frequency of 440Hz, has an A note an octave *above* it at 880Hz, and an octave above that at 1760 Hz, and so one. There is also an A an octave *below* 440Hz at 220Hz, and 110Hz, and so on. Speech researchers settled on the slope of –12dB per octave as an *idealized* amplitude drop off in higher frequencies (Ni Chisade and Gobl in Handbook of phonetic sciences), though in reality, the slope in normal voices varies tremendously and does not decrease evenly across the spectrum (Kreiman et al., 2007). Let's use the –12dB/octave as a guide for spectral slope, for the moment, to illustrate what the overall spectrum would look like in different scenarios. Let's assume that the average male F_0 is 130 Hz, and 200 Hz for the average female voice.[3] How would the glottal spectra differ assuming a roll-off of 12 dB/octave?

[3] Fundamental frequency is *very* variable.

TWO SPECTRA, ONE FOR MALE F0=130, FEMALE F0=200

A few things to notice from the image above are the energy at the various frequencies, and their decreasing amplitude. In the spectrum on the left representing the average male voice, the F_0 (130Hz) has an amplitude of 60dB. The higher-level harmonics are at whole integer multiples of 130: 260Hz (130 x 2), 390Hz, 520Hz, 650Hz… An octave *above* F_0 is 130 x 2, or the 260Hz harmonic (here H2, or *second harmonic*) of the spectrum. Given the spectral slope we specified (-12dB/octave), H2 has an amplitude of 48dB (60 -12). The next harmonic we can predict the amplitude for is 520Hz (260Hz x 2), or H4, which will be 36dB (48-12). Likewise, H8 (H4 x 2), H16 (H8 x 2) will progressively drop in amplitude by 12dB each time. The very same operation can be used with the glottal spectrum on the right, where the 400Hz component (H2) would have an amplitude of 37dB, and H4 25dB, and so on.

4.4 The quality of voices

Vocal *quality* is a nebulous term that we might have a sense about without a precise definition. It captures the differences between voices that are perceptually salient, the identifying features that lets you know who is talking. It's the same thing that might distinguish different musical instruments from one another, but here applied to voices. It's rare that someone cannot tell a piano from an acoustic guitar even if they're playing the exact same note, that is, we're so familiar with how a piano and guitar sound that they're identifiable with just one note. Similarly, we are so accustomed to how voices sound that we can tell them apart instantly (more or less)

especially if we're familiar with them—how often do you confuse the voice of your friend with the voice of your mom?

But what is it about voices that allow us to do this? Especially when the two voices might have the same, or very similar, fundamental frequencies. Say for example two people have the exact same fundamental frequency. This would mean that the harmonic content of their glottal spectrum is essentially the same (remember higher harmonics are integer multiples of F_0), yet the voices *sound* different. This reflects the differences in the *downstream* filtering of the sound coming from the vocal fold vibration. Vocal tracts are different with cavities of varying lengths and cross-sectional areas. Sinus cavities, which also resonate, are not identical between people and they have varying amounts of gunk in them affecting the sounds coming from the glottis. These differences in oral cavity dimensions will filter the glottal spectrum in different ways, resulting in a vocal *signature* of sorts. For example, vocal tract lengths (from glottis to lips) of males and females are quite similar up until around age 8 (around 13cm), when they start to diverge. By age 20, they differ by about 2cm, with males vocal tract lengths around 17.5cm and females around 15.5cm on average (Story et al., 2018). Such a difference in vocal tract length has consequences for what areas of the glottal spectrum are amplified and dampened.

If we go back to our tube models in Chapter 3, we can measure the effect on a neutral vowel like schwa (or the sound "uhh"). If the vocal tract is modeled like a tube that's closed at the glottis end and open at the lips, we can calculate its resonances and the areas of the spectrum that are amplified.

<u>For the average male vocal tract (17.5cm):</u>

$F_1 = v/4L = 343 \text{ m/s} / 4(17.5) = 1500 \text{ Hz}$

<u>For the average male vocal tract (15.5cm):</u>

$F_1 = v/4L = 343 \text{ m/s} / 4(15.5) = 1330 \text{ Hz}$

The difference in the first resonance of the same vowel in voices different sized vocal tracts but with the same glottal spectrum contribute to the quality difference between the voices.

Voice quality can also reflect the ways vocal folds vibrate in an individual. When vocal folds are taut, causing them to be pressed when adducted, P_{sub} must be high in order overcome the tension to initiate phonation. This **hyperadduction** is often described as *harsh* or *tense*. The opposite hyperadduction is **hypoadduction**, which results from vocal folds not fully adducting. With the loose tension of the hypoadducted vocal folds, the P_{sub} does not find much resistance to begin phonation. The hypoadducted vocal folds might allow air to escape without being converted to acoustic energy. The loose tension can create a narrow channel between the vocal folds adding turbulent noise to the glottal spectrum. The hypoadducted voice is often described as *breathy* or *hoarse*.

Vocal fold thickness which varies from individual to individual, can also affect how we sound, in particular the slope of the glottal spectrum. The thicker the surface of the vocal folds, the shallower the slope, meaning there is louder acoustic energy in the higher harmonics of the spectrum. Whereas thin vocal folds lead to a greater volume velocity of air flow (Zhang, 2016). Another factor leading to individual voice signatures relates to just how periodic vocal fold vibration might be. Vocal folds, like any tissue, are not uniform in thickness, density, elasticity, etc., like an engineered plastic or metal, and so they're subject to variability in their movement. Small fluctuations in their movement results in variation in frequency and amplitude of the acoustic energy being produced by their vibration. Recall our discussion of simple harmonic motion (Ch.2), where the vibration of a tuning fork produces a nearly perfect amplitude wave where every cycle of compression and rarefaction is more or less identical. Well, vocal folds are

not a tuning fork, so every cycle of opening and closing will not be identical in terms of their periodicity as well as the amplitude of displacement of air passing through. These variations have special names: **jitter** refers to frequency variation, and **shimmer** is amplitude variation. Each one of us has varying amounts of jitter and shimmer which further adds to the quality of our voices.

4.5 A variety of voices

Phonating comes in a variety of *flavours* that we're very familiar with. The falsetto of a pop singer, the "vocal fry" of podcaster, or the shouts of children on a playground are all varieties of phonation that we call **registers**. Every phonation register has its own vocal fold configuration and requirements in terms of the volume velocity of air travelling through the glottis. Below I'll outline some common registers (in order of their fundamental frequency range), the volume velocities of air and subglottal pressures (P_{sub}) involved, and how exactly the vocal folds need to be situated in order to produce the effect.

> ***Think about it!***
> *Vocal fry* is a phonation register often associated with young people, and more specifically, young women. A 2011 article in the journal *Science* claims that the phonation style was "creeping in" to American Speech, becoming a "language fad." While some (older) people find it annoying, lowering your fundamental frequency is a very common phenomenon that *everyone*, men and women, participates in when speaking phrases or sentences. As the air from your lungs slows down and decreases in pressure over the course of your utterance, your fundamental frequency naturally decreases below your *normal* modal phonation level. It signals to the listener that your sentence is coming to an end.

1. **Pulsed** (*Audiobank* sample): The lowest fundamental frequencies produced by speakers is variously called *pulsed voice*, *creaky voice,* and more popularly called "vocal fry" or "glottal fry" (keep in mind that these last two aren't technical terms). The F_0 in the pulsed register is on average around 50Hz (Blomgren et al. 1998). The low fundamental

frequency results from the short and relatively thick vocal folds. The vocal folds also may contact the false vocal folds, thereby increasing the mass of the vibrating length (remember, increased mass will decrease the fundamental frequency (Blomgren et al. 1998). The vocal folds are mostly adducted and a lower (highly variable) subglottal pressure is required to separate them (around 5cm H_2O) (Venkatraman and Sivasankar, 2018). The opening and closing movement of the vocal folds is quite different from the more typical modal register (described below). Instead of fully opening and closing, the anterior (front) portion may open and oscillate along the length many times before fully adducting again. This results in a *tapping* or *sputtering* quality when sustained. The glottal spectrum of the pulsed register reflects a low F_0, and subsequent gradual decreasing amplitude.

2. **Modal** (*Audiobank* sample)**:** The modal register is also called **voiced** phonation or *normal* voice. The term "modal" refers to the regular oscillation of the vocal folds, which is the natural function (or mode) of the vocal folds. It's the most used register in speaking, used producing vowels and the voiced series of consonants like *b,d,g,v,z,*or *th* (as in "the"). The modal register generates the widest fundamental frequencies (and amplitudes) used in speech, with an average of around 100-140Hz for males and 175-240Hz for women (Krook, 1988). For voicing the glottis is closed, with the vocal folds adducted with varying degrees of tension. The vocal folds are shorter in modal phonation relative to other registers, and the entire length of the vocal folds vibrate once air from the lungs has achieved the threshold pressure, which is typically around 7-8 cm H_2O (Krook, 1998). The volume velocity of air required for modal voicing is at least 50cm^3/sec, and typically ranges 70-200 cm^3/sec. For higher frequency voicing (greater than 130Hz), the volume velocities are higher, as are the phonation threshold pressures.

3. **Falsetto** (*Audiobank* sample): For falsetto phonation, where an individual might artificially use a *very* high-fundamental frequency voice in speaking or singing (around 300-600Hz in men and 500-1000Hz in women), the vocal folds are very stiff and elongated (and thin, like a stretched rubberband) due to tension exerted by the cricothyroid muscle. The tautness of the vocal folds prevents the glottis from being completely closed. Due to the stiffness of the folds, the pressure required to set them into motion is higher than normal modal voicing at around 7cm H_2O, with a volume velocity of air greater than 70cm^3/sec. The *thin* quality of falsetto phonation results from the limited vibration of the vocal folds, the entire length of the vocal folds doesn't vibrate in as complex a manner as they do in the modal phonation register.

All these phonation types result in a glottal spectrum with harmonics that are amplified (or dampened) by the oral cavity. But there are registers where sound is generated at the glottis and *does not* result in a complex periodic vibration of the vocal folds. While these sounds aren't typically called *phonation*, they nonetheless are important for our understanding of the variety of sounds that are filtered downstream by the oral cavity.

4. **Voiceless** (*Audiobank* sample): Breathing through an open glottis results in the *voiceless* phonation type that is heard in sounds *p,t,k,f,s,sh,h*. There are two types of voiceless phonation, the first with turbulent, and consequently noisy air flow (also called **breath**), and the other with laminar, or noiseless flow (also called **nil**). In both cases, the vocal folds are wide open. For breath phonation, volume velocities of air are greater than 200-300 cm^3/second, while for nil phonation, the velocities are below that level. It's not uncommon to see very high velocities (around 1000 cm^3/second) for breath phonation as in the *h* sound at the beginnings of words, with lower flow rates in nil-phonation in sounds like *f, s, sh*, where the noisiness is not generated at the glottis but upstream in mouth (Catford, 1977).

5. **Whisper** (*Audiobank* sample): The whisper register does not involve vocal fold vibration, but rather the phonation comes from the noise generated by the high-velocity air flow, around 200-300cm^3/second (Monoson et al., 1984), through a narrow opening of the glottis (about 25% of the length of the vocal folds). The volume velocity of air through this relatively small channel results in a flow that is highly turbulent and noisy. Unlike voiced phonation, there is energy *across* the spectrum, with amplitudes varying from frequency to frequency (aperiodic), which is then filtered by the oral cavity like vowels.

These phonation types are by no means exhaustive (e.g., *breathy voice, whispery voice, murmur, double voice, ventricular creak*, etc.), but serve as a good starting point in understanding the variety of ways that the glottal opening can shape the air escaping your lungs in acoustically significant ways.

4.6 Disordered phonation

"Normal" phonation is a rather subjective term, but speech-language pathologists agree on some fundamental aspects of the normally functioning glottis and the acoustic properties of the sounds it produces. We've described above some of the factors that contribute to variation in phonation between people including structural differences in the shape, length, and tension of the vocal folds, but there is a range for certain acoustic characteristics that typify a normal voice. Apart from average F_0 (around 100Hz in male adults and 200Hz in female adults), adults usually have a F_0 range of about 2.5-3 octaves and can maintain phonation for around 15-25 seconds. Normal

voices typically have less than 1% variation in the periodicity (jitter) of their vocal fold vibration.

> ***Think about it!***
> You're watching the latest thriller on TV, anticipation building when suddenly the main character screams in fright. You audibly gasp! But what *is* that sound? It's technically an *inspiratory* phonation, which is basically the reverse of normal phonation in terms of the flow of air. This happens when inspiratory air flow (increased volume of the thoracic cavity leading to low P_{sub}) sets the vocal folds, which are not full abducted, into motion. This can happen because of glottic stenosis, or the narrowing of the glottal opening which is often a result of intubation after injury. But when you gasp when it's a *paralinguistic* sound that let's others know you're scared!

The image below shows five cycles of vocal fold vibration from an individual with disordered phonation. To calculate the *jitter* you will need to calculate the average cycle-to-cycle variation and then divide it by the average duration of the cycles. The jitter for this individual is greater than 1%, which is often diagnostic of a phonation disorder.

IMAGE OF 5 GLOTTAL CYLCES WITH T1-5 BOUNDARIES WITH DURATIONS

[(T2-T1) + (T3-T2) + (T4-T3) + (T5-T4)]/5 = mean jitter

(T1 + T2 + T3 + T4 + T5)/5 = mean period

Percent jitter = (mean jitter/mean period) x 100

All normal voices have some degree of noise, or aperiodic acoustic energy in the glottal spectrum resulting from obstructions to the airflow. This noise is usually very low amplitude, that is, the signal-to-noise ratio (**SNR**) is high, which means that the periodic energy from the glottis is louder than the noise (Zemlin, 1998).

Disordered phonation can refer to a host of deviations from normal voice quality, fundamental frequency, or amplitude, but in lay terminology is often captured with descriptions like *shrill*, *raspy,* or *hoarse,* etc. These aren't diagnostic terms for disordered phonation but can

allow clinicians to follow up with individuals who might present with voice quality conditions that are noticeable. For example, a breathy voice indicates that the vocal folds do not close completely (but they're close enough to have vibration initiated) and as result, there is continuous air flow through the glottis during vocal fold vibration. Air flow through the narrow channel of the glottis is turbulent and adds noise to the higher frequencies of the glottal spectrum. A hoarse voice is often the result of irritated or swollen vocal folds, which affect the periodicity of their vibration. Aperiodicity is essentially noise that is added to the periodicity of the vocal fold vibration and in hoarse voices the noise is in the lower frequencies. Clinicians might use a harmonics-to-noise measure to evaluate the extent of breathiness or hoarseness in an individual's voice, which might indicate an obstruction in airflow (from a growth on the vocal folds) or neurological issues leading to aperiodicity in the vocal fold movement.

But why might phonation become disordered in the first place? Perhaps you've experienced the feeling of losing your voice (not metaphorically, but literally!). It often accompanies some sort of sickness like a cold or the flu which can result in a swelling of larynx that disrupts normal phonation. Infections like this also tend to cause excess mucous production, which then coats the glottis, affecting the periodicity of the vocal fold vibration. **Vocal hygiene** can also contribute to phonation that deviates from normal function. This refers to a variety of external factors that that affect the vocal folds, like smoking and poor hydration, as well as activities that stress the glottis, like screaming or even excessive throat clearing. But while viral and bacterial infection, and poor habits are everyday sources of disordered phonation, there are perhaps more significant reasons why phonation deviates from normal functioning. Neurological disorders can also affect phonation.

Progressive neurological disorders (like Parkinson's Disease or Amyotrophic Lateral Sclerosis), or neurological disorders that typically involve increasing deterioration in function over time, can result in deviations from normal phonation. Parkinson's Disease, which affects the dopamine (a neurotransmitter molecule that plays an important role in body movement and the feeling of pleasure) production. The typical symptoms of reduction in dopamine resulting from Parkinson's Disease include tremors, slowness in initiating movements, and muscle rigidity. Phonation may be one of the first signs of Parkinson's disease appearing before other speech deficits like articulatory and fluency impairments (Ho et al., 1998). Some of the more common Parkinson's related phonation problems is reduced amplitude of voice, higher F_0, and decreased overall range of F_0 (*Audiobank* sample). Why might the vocal folds be targeted by Parkinson's? There may be a relationship between the overall effects of Parkinson's on the motor functioning of muscular structures in the body and the vocal folds. One study found that patients with more rigidity (in their limbs) on a particular side of their body likewise had more rigidity on the same side of the vocal folds. Rigidity in the vocal folds leads to often incomplete closure, phonation weakness, and breathiness (Moro-Velazquez et al., 2021).

There are also neurological disorders that target the nerves of the larynx. Spasmodic dysphonia (also called *laryngeal* dystonia) is a neurological condition that causes spasms (sudden, involuntary movements) in the muscles of vocal folds. Individuals with spasomodic dysphonia (*Audiobank* sample) have periods of phonation *breaks* with a voice that sounds strained or breathy. Sometimes individuals with this dysphonia have a vocal tremor because of these vocal fold spasms. There are different varieties of this disorder, targeting either the adducting or abducting motions of the vocal folds. Adductor spasmodic dysphonia is the more common of the two, which causes the vocal folds to snap closed and stiffen, leaving the voice

sounding strained. With abductor spasmodic dysphonia the vocal folds stay open, which can affect how readily the vocal folds vibrate. Vocal fold vibration is difficult to initiate because of their open position and when they do vibrate the amplitude is reduced. The voice of individuals with abductor spasmodic dysphonia ends up sounding weak and breathy.

Summary

Phonation is the consequence of the pressure differential between sub-glottal and supra-glottal air. Adducted vocal folds are set into motion when phonation threshold is achieved, causing fast moving air to flow through the glottal opening. Vocal folds then recoil back to an adducted state due to their natural elasticity as well as the Bernoulli effect. The process then repeats. Vocal fold vibration results in a *glottal spectrum*, or acoustic energy in various frequencies determined by the primary vibration (fundamental frequency). Importantly, the spectrum reflects a complex periodic sound, where harmonics are related to the fundamental frequency. Phonation can come in different varieties, or *registers*, depending on the state of the vocal folds, how taut or loose they are and whether they are fully adducted or not. Phonation is often targeted by neurological disorders that affect muscle movement of the larynx.

Important definitions/concepts

Abduction: vocal folds moving apart, opening the glottis

***Adduction*:** vocal folds coming together, closing the glottis

***Anterior*:** in anatomical directions, the front of the structure in question

***Bernoulli Effect*:** the physical principle which states that an increase in the velocity of air results in a decrease in pressure

Falsetto: the high fundamental frequency used in some kinds of singing or vocal *affect,* and resulting from thin and taut vocal folds

***Glottal source*:** acoustic energy generated by the vibrating vocal folds

***Glottal spectrum*:** the spectrum of acoustic energy reflecting the complex periodic vibration of the vocal folds

***Inferior*:** in anatomical directions, the lower region of the structure in question

***Jitter*:** variation in the frequency of vocal fold vibration during voiced phonation

Modal: a distinct pattern of vibration of the vocal folds which is periodic and otherwise regular; contrast with *amodal* vibration, which is irregular.

***Myoelastic aerodynamic theory of phonation*:** the *basic* theory explaining how and why vocal folds vibrate. The myoelastic theory suggests that adducted vocal folds are blown apart by high pressure sub-glottal air and once open recoil back to adducted position due to their elasticity in addition to the Bernoulli Effect

***Octave*:** the relationship between two frequencies such that the higher frequency is twice the lower frequency. For example, 440Hz is an octave *higher* than 220Hz, or 220Hz is an octave *lower* than 440Hz.

***Oscillation*:** the back-and-forth movement of the vocal fold vibration

***Phonation*:** typically refers to the acoustic energy resulting from vocal fold vibration (also called *voicing*), but in some literature is expanded to mean any acoustic energy (even aperiodic noise, such as *whisper*) that can be filtered by downstream cavities in the head

***Phonation threshold*:** the sub-glottal pressure required to blow open the adducted vocal folds

***Posterior*:** In anatomical directions, the back portion of the structure in question

***Pulsed*:** A voicing register with a very low fundamental frequency, sometimes called *creaky voice,* or *vocal fry*, resulting from compressed and slack vocal folds.

***Registers*:** Characteristic types of phonation; exemplified by pulsed, modal, falsetto, voiceless, and whisper phonation types

***Shimmer*:** Variation in the amplitude (or displacement) of vocal fold vibration during voiced phonation

***Single-mass model*:** Model of vocal-fold vibration where each fold is modeled as a single and independently oscillating mass. It extends the myoelastic theory by introducing the aerodynamic contribution of supraglottal pressures

***SNR*:** Signal-to-noise ratio, an index of how noisy the speech (or any audio) signal is. The lower the SNR, the noisier the signal. For example, 0dB SNR means equal amplitudes of signal and noise.

***Source*:** The acoustic signal that is ultimately filtered structures downstream. For example, vocal fold vibration is a source that is filtered by the pharyngeal, oral, and sinus cavities. Turbulent air through an open glottis (that does not result in vocal fold vibration) is also a source sound.

***Spectral slope/ roll-off*:** The rate at which the amplitude of the harmonics of the glottal source spectrum diminishes. The theoretical roll-off for the glottal spectrum is -12dB/octave.

***Sub-glottal pressure*:** air pressure below the glottis; in order to initiate phonation it must be greater than the supra-glottal pressure in the pharynx and mouth

***Superior*:** in anatomical directions, the top of the structure in question

***Vocal hygiene*:** the good oral habits for healthy vocal folds like hydration, and not smoking or shouting excessively

***Voiceless*:** a phonation register characterized by relatively open (abducted) vocal folds, allowing either turbulent air flow (*breath*) or laminar air flow (*nil*)

***Voicing*:** a phonation register characterized by closed (adducted) vocal folds, which are set into vibration by the pressure differential between sub-glottal and supra-glottal cavities; in some literature *voicing* is synonymous with *phonation*

***Whisper*:** a phonation register characterized by noise of high-velocity air flow through a narrow opening in the glottis with no vocal fold vibration

---

Praat exercise: **Synthesizing a glottal spectrum**

What would your vibrating vocal folds sound like if you didn't have a head? *Praat* allows you to create a glottal spectrum, which we've idealized above (4.3), by either starting from scratch, or "reverse engineering" the glottal waveform from a vowel sound you've recorded. In this exercise you'll create the glottal spectrum from scratch using a **PointProcess** object. Here is the process:

1. First create a pitch tier (New > Tiers > Create PitchTier…), you can name the pitch tier "GlottalSource" and give it a duration of 1s.
2. Edit the PitchTier object. You'll see an empty field in which you can add "pitch points," specifying the fundamental frequency of your source spectrum.

-For this exercise let's create a glottal spectrum with a jitter of 0. This would mean that the frequency of vocal fold vibration does not change over the 1s of the pitch tier object.

-Pick a fundamental frequency, say 125Hz. PitchTier GlottalSource > View & Edit > PitchTier > Add point at… [Time = 0, Frequency (Hz) =125]. Now do the same thing, but add a point at Time = 1, Frequency = 125. Now you have a pitch tier object with an even 125Hz fundamental frequency throughout the object.

3. From the Objects window, select the PitchTier GlottalSource object > Synthesize > To Sound (phonation). This pulls up a form with lots of options, such as "Adaptation factor" and "Power". You can read *Praat* documentation about these options if you want to learn more about the algorithm that generates the phonation waveform, but for now, leave them be and click "OK"

4. You now have a *Sound* object, which is the phonation waveform. View & Edit the sound.

5. Zoom in on the waveform and observe its shape. It is *not* a sine wave (thankfully, because our vocal folds are not tuning forks!)

6. Select the entire waveform, then Spectrogram > View Spectral slice. You should see the individual harmonics of the phonation waveform synthesized from the PitchTier object by *Praat.* Zoom in on the first few harmonics, and for a sanity check measure the

frequencies of the first three peaks. They should be around 125Hz, 250Hz, and 375Hz, or integer multiples of 125Hz!

7. Can you calculate the roll off based on this synthesized spectral slice?

Practice Problems

1. In the glottal spectrum below, what is the amplitude (dB) of the 8th harmonic given a roll-off of -13dB/octave?

IMAGE OF A GLOTTAL SPECTRUM (FREQ HZ x AMP DB) WITH 10 HARMONICS, GIVING LABELS FOR ONLY THE SECOND HARMONIC

2. Calculate the percent jitter of the glottal cycles below. Would this glottal waveform be considered disordered phonation according to the 1% threshold?

IMAGE OF 10 GLOTTAL PULSES OF VARYING DURATION

References

Catford, J.C. (1977). *Fundamental Problems in Phonetics*, Bloomington: Indiana University Press.

Dedo, H. H., & Dunker, E. (1967). Husson's theory: An experimental analysis of his research data and conclusions. *Archives of Otolaryngology*, 85(3), 303-313.

Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L., & Gates, S. (1998). Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural Neurology*, *11*(3), 131-137.

Monoson, P., & Zemlin, W. R. (1984). Quantitative study of whisper. *Folia Phoniatrica et Logopaedica*, *36*(2), 53-65.

Moro-Velazquez, L., Gomez-Garcia, J. A., Arias-Londoño, J. D., Dehak, N., & Godino-Llorente, J. I. (2021). Advances in Parkinson's disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects. *Biomedical Signal Processing and Control*, *66*, 102418.

Story, B. H., Vorperian, H. K., Bunton, K., & Durtschi, R. B. (2018). An age-dependent vocal tract model for males and females based on anatomic measurements. *The Journal of the Acoustical Society of America*, *143*(5), 3079-3102.

Venkatraman, A., & Sivasankar, M. P. (2018). Continuous vocal fry simulated in laboratory subjects: A preliminary report on voice production and listener ratings. *American Journal of Speech-Language Pathology*, 27(4), 1539-1545.

Zemlin W. R. (1998). *Speech and hearing science: anatomy and physiology* (4th ed.). Allyn and Bacon.

Zhang, Z. (2016). Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. *The Journal of the Acoustical Society of America*, *139*(4), 1493-1507.