Music and Medicine
2(2) 104-108

DOI: 10.1177/1943862110365881
http://mmd.sagepub.com
SAGE

The Human Voice: Evolution and Performance

Michael S. Benninger, MD

Abstract

From the first sounds of an infant until the end of life, the human voice is the most important means of human communication. The development of voice and speech may be the most important evolutionary development in the division of humans from other animals. The voice can project all of human emotions from our greatest joys to our deepest sorrows. The ability to sing was fundamental to the development of music and culture. This article will describe the significant evolutionary adaptations of the larynx and pharynx that resulted in the development of the human voice. This unique ability to communicate led to the development of speech and song, the starting point to the creation of music. How the voice works, the correlation of the voice to other musical instruments, and the interaction of other neurological functions will be presented.

Keywords

voice, evolution, music, phonation, singing

What if you had no voice? What if you could not speak, could not shout, or could not sing? The human body has remarkable attributes that have allowed us to function in even the most inhospitable environments, while still providing stimulation and joy and satisfaction. But what are those unique characteristics that have distinguished humans most from other animals? Arguably, there have been a few major evolutionary developments that have allowed human beings to become the preeminent species on the planet. Upright walking and the opposable thumb are frequently cited as key steps in the evolutionary uniqueness (Abitbol, 2005). Yet, they may not have been the key factors in the development of civilization and ultimately the development of culture and music. Arguably, it is voice and speech that have allowed us to organize, to establish communities, and eventually to develop music.

The human larynx and pharynx are unique in the animal kingdom. There are key aspects of these anatomic structures that have resulted in the ability to produce language and the flexibility to produce song. The layered structure of the vocal fold is also unique among animals and is the reason that the vibratory characteristics of the human larynx can permit the volume to shout and the flexibility to sing an aria and can still allow production of a soothing song to a sleeping infant. This article will briefly describe the anatomy and physiology of the voice, the remarkable evolutionary changes that occurred to allow for the development of speech and singing, and how the uniqueness of the human voice was critical to the development of music.

Laryngeal Anatomy and Physiology

Musical instruments have three major components: There is something that activates the sound (plucking a string, striking a piano key), a vibrator to produce the desired frequencies (the string of a violin, the reed of wind instruments), and a resonator (the body of the instrument that gives each instrument its unique sound). The human body is similar. The lungs serve as the activator of the sound, the vocal folds vibrate to produce the desired frequencies, and the resonating cavities of the pharynx, oral cavity, nose, sinuses, and chest modify or amplify the sounds. Unlike other musical instruments, however, the production of voice and singing is much more complex than just these three key components. It is an interaction of the vocal tract, the abdomen and diaphragm, the musculoskeletal system, and the psychoneurological system. All of these need to be functioning properly and in a coordinated fashion to produce the desired sound. In fact, most muscles of the body in some way are involved in the production of the singing voice.

It is also interesting that the larynx is in the middle point of the central highway of the body, in the neck. Through this relatively narrow structure, the brain is connected to the rest of the body. The nerve and blood supply to and from the brain lie adjacent to the larynx. The airway and upper digestive system are integrally involved with the function of the vocal folds.

The primary structures for the production of voice are the vocal folds. These relatively small structures (1.5 to 2.3 mm in length) require 22 muscles to allow them to function in the

The Cleveland Clinic, OH, USA

Corresponding Author:
Michael S. Benninger, Head and Neck Institute, The Cleveland Clinic, 9500 Euclid Avenue, A-71, Cleveland, OH 44195
Email: benninm@ccf.org

precise fashion that is necessary to produce voice. The principle joint (thyroarytenoid joint) that allows each vocal fold to move is one of the most complex joints in the body, providing three degrees of movement: rocking, gliding, and rotation. These are adjusted by minute changes in each of the laryngeal muscles. Phonation is the term used to describe the production of voice with vocal fold modification determining the frequency and is the major determinant of pitch.

The ability to produce such a large range of frequencies requires vocal fold vibration to occur more quickly than muscle contraction. To accomplish this, aerodynamic forces need to be modified to allow for rapid vibration in a coordinated fashion. The primary model believed to account for the vibration is the Myoelastic Aerodynamic Theory of Phonation (Van de Berg, 1958). The vibration of the vocal folds depends on two components. Myoelastic refers to neuromuscular control of the vocal folds and the elastic properties needed for phonation, while the aerodynamic properties refer to the airflow and fluid dynamics and the effect of Bernoulli's principle on that movement (Jiang, 2006).

The basic principles of the production of voice begin with tension being applied to the vocal folds to bring them into a contact position or adduction. As airflow passes between them primarily with expiration, the vocal folds are set into motion. The air pressure in the area below the vocal folds, or subglottic pressure, increases to overcome the tension of the vocal folds holding them together, essentially pushing or blowing them apart. As the air passes through the vocal folds, the subglottic pressure drops quickly and the pressure of the muscular tension of the vocal folds becomes greater than the subglottic pressure and the vocal folds will again come back together. Participating in this approximation is the Bernoulli principle. As air passes through the narrowed glottis, the reduced pressure draws the vocal folds together (Jiang, 2006). When the subglottic pressure builds up again, the vocal folds again are pushed apart and the cycle repeats itself. This can occur multiple times a second, allowing for the ability of the human larynx to produce vibrations or a frequency well over 1,000 cycles per second in some people.

Therefore, there are a number of important factors that help to determine frequency or pitch. They include the mass of the vocal fold, the tension of the vocal folds, and the level of subglottic pressure. Similar to a piano or guitar string, and with all other things being equal, a larger or heavier vocal fold will vibrate more slowly than a smaller or thinner vocal fold. This is one of the factors that explains the higher range of female compared with male voices or the lower pitched voice of a smoker whose vocal folds are edematous. It is also the key reason that patients with masses on their vocal folds, such as vocal nodules, have a reduction in the upper portion of their singing range while their speaking voices sound huskier or lower pitched. The effect of tension of the vocal folds is also similar to a string instrument. The tighter the string is strung, the higher the frequency. In the larynx, the more tightly the muscle contracts, the higher the frequency is produced. Lastly, the higher the subglottic pressure that is created, the more rapidly will the

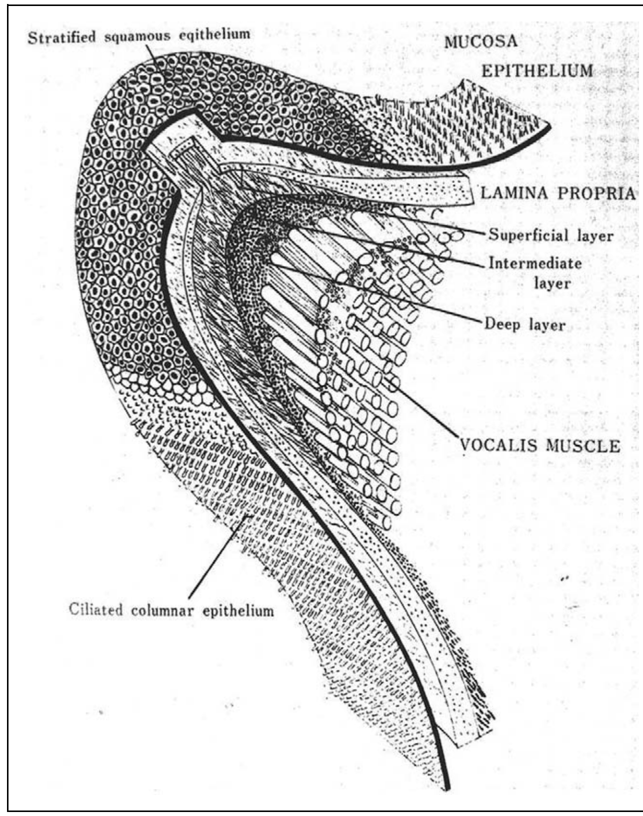

Figure 1. The layered structure of the human vocal fold with an epithelium, three layers of lamina propria, and vocalis muscle.
Source: Hirano (1975, p. 241).

cycles occur, elevating pitch (Coulton, 1994). Pitch is therefore increased with decreased mass, increased vocal fold tension, or increased subglottic pressure. Pitch is decreased with increased vocal fold mass, decreased vocal fold tension, or decreased subglottic pressure. It can be speculated that aspects of human emotion that alter muscle tension may play a role in pitch and pitch control.

For the vocal folds to be able to vibrate in such a rapid and complex fashion, they need to be pliable and flexible. In the early 1970s, Dr. Minoru Hirano described the layered structure of the vocal folds as consisting of a body (the thyroarytenoid or vocalis muscle) and a cover, which is the tissue that covers the muscle (Hirano, 1974, 1975; Hirano et al., 1973; see Figure 1). The anatomy of the cover is complex in and of itself, with a superficial epithelium, two intermediate layers of the lamina propria, and a thick deep layer or thyroid ligament (see Figure 1). Each of these layers has the ability to vibrate independently (Hirano, 1974) or as a unit, giving remarkable flexibility to the voice. Differences between modal tone, chest voice phonation, or falsetto may be explainable by the coupling or independent movement of these layers (Coulton, 1994).

The deep layer of the lamina propria or vocalis ligament is an interesting structure. It is not present at birth and although there is thickening beginning by age 8, the lamina propria is not fully developed until the age of 11 or 12 (Hartnick, Rehbar, & Prasad, 2005; Prades et al., 2009). Although a pediatric larynx

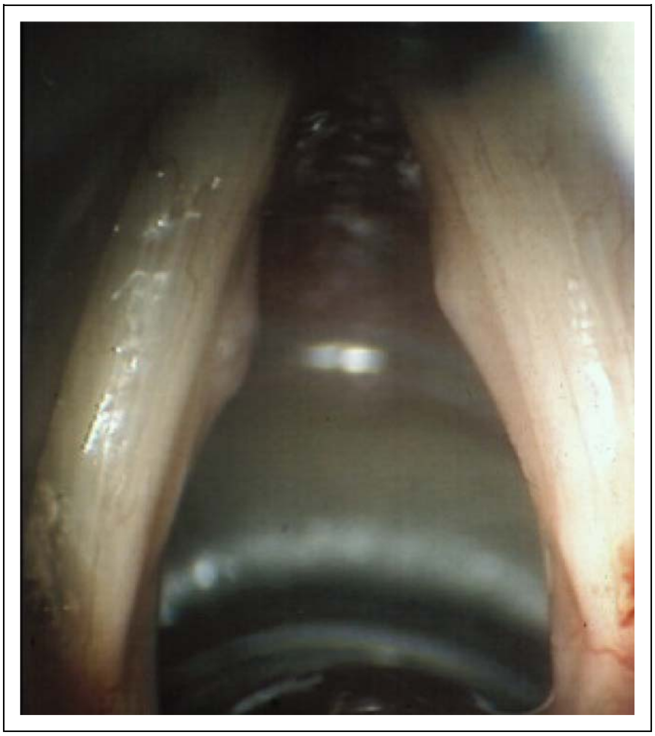

Figure 2. Vocal fold nodules in a 32-year-old female singer.

is less likely to develop dysphonia (abnormal voice), the pediatric larynx is less able to perform fine vocal articulations than an adult (Prades et al., 2009). Another interesting feature of the human lamina propria is that no other animal has a deep layer of the lamina propria, which may in part account for the distinctions between the phonation of humans and other species. The changes in voice that occur with age may also be in part related to a reduction in the thickness of the lamina propria and density of the epithelium (Hartnick et al., 2005).

In most cases of vocal injury, the area involved is confined to the epithelium and to the superficial layers of the lamina propria. The classic examples of this are vocal nodules (see Figure 2) and most vocal fold polyps. In the case of nodules, recovery is expected with changes in the intensity, quantity, or technique of voice use (Benninger & Jacobson, 1995). Injury or surgery confined to the epithelium or superficial layer of the lamina propria will heal with essentially no effect on vocal fold function, since the fibers of the lamina propria will reestablish in an orientation parallel to the epithelium. When injury or surgery involves both the epithelium and the deep layer of the lamina propria, then the scar can develop perpendicular to the epithelium and result in stiffness and reduction of vocal wave (Benninger, 2000; Benninger et al., 1996). This in turn can result in dramatic, and at times permanent, changes of voice. The characteristics of the ultrastructure of the human vocal fold with a vocalis ligament not only may result in the unique qualities of the human voice but also may predispose to greater risk of permanent changes in voice after injury.

The Uniqueness of the Human Voice

The three unique functions of the human larynx are for protection, respiration, and phonation (Tucker, 1994). Evolutionarily and functionally, protection and airway are the primary functions. The protective effect is the reason that the vocal folds can go into spasm (laryngospasm) or cause aggressive coughing if there is some aspiration. In humans, the larynx serves as a primary organ for communication and, phylogenetically, this is the most recent or youngest function (Tucker, 1994). Many animal species have larynges that are grossly similar to humans, minus the vocalis ligament, and many have the capacity to create various pitches, often at very low or high frequencies. None of them have been able to develop the complexities of the human voice, a language, or the ability to sing other than in the most simplistic way.

What has occurred that has led to the unique ability of the human voice? Perhaps the most significant evolutionary development that has led to the ability for humans to develop speech and language has been the descent of the larynx in the neck. In all other animal species, the larynx during normal respiration is situated high in the neck, between cervical vertebrae 1 and 3 (Laitman, Van De Water, & Noden, 1995). This laryngeal position correspondingly results in higher positions of the tongue, the epiglottis, the hyoid bone, and the pharyngeal constrictors. The tongue lies almost completely in the oral cavity and not in the pharynx. In such a position, the epiglottis abuts and in many cases overlaps the uvula and soft palate, allowing the larynx to open directly into the nasopharynx and resulting in the ability for the lungs to communicate through the laryngeal airway directly with the nose (Laitman et al., 1995). Food or fluids can pass on either side of the interlocked palate and larynx directly into the pyriform sinuses and esophagus, resulting in patency of the laryngeal airway while the animal can simultaneously eat and swallow (Laitman & Reidenberg, 1993). Infants have a similar pharyngolaryngeal configuration, which allows the infant to suckle and breathe simultaneously and is one of the reasons that infants are primarily nasal breathers (see Figure 3). This high laryngeal position, however, significantly inhibits the ability of infants and small children to modify sounds and limits the array of sounds that can be produced (Laitman et al., 1995).

There are clearly distinct teleological advantages to this separation of swallowing and breathing functions. Since smell function is very refined in many animal species and is used as a source of protection from predators, the ability to breathe through the nose, and therefore smell, while eating would afford some survival benefit. This also would allow the animal to run while still eating. Lastly, this configuration dramatically reduces the risk of aspiration (Laitman & Reidenberg, 1993). The high neck position of the larynx, however, strongly adversely affects the ability to modify sounds. As described by Laitman,

> the high position of the larynx greatly restricts the supralaryngeal portion of the pharynx available to modify the initial, or fundamental sounds produced by the vocal folds. Thus an individual with a larynx situated high in the neck, such as found in a

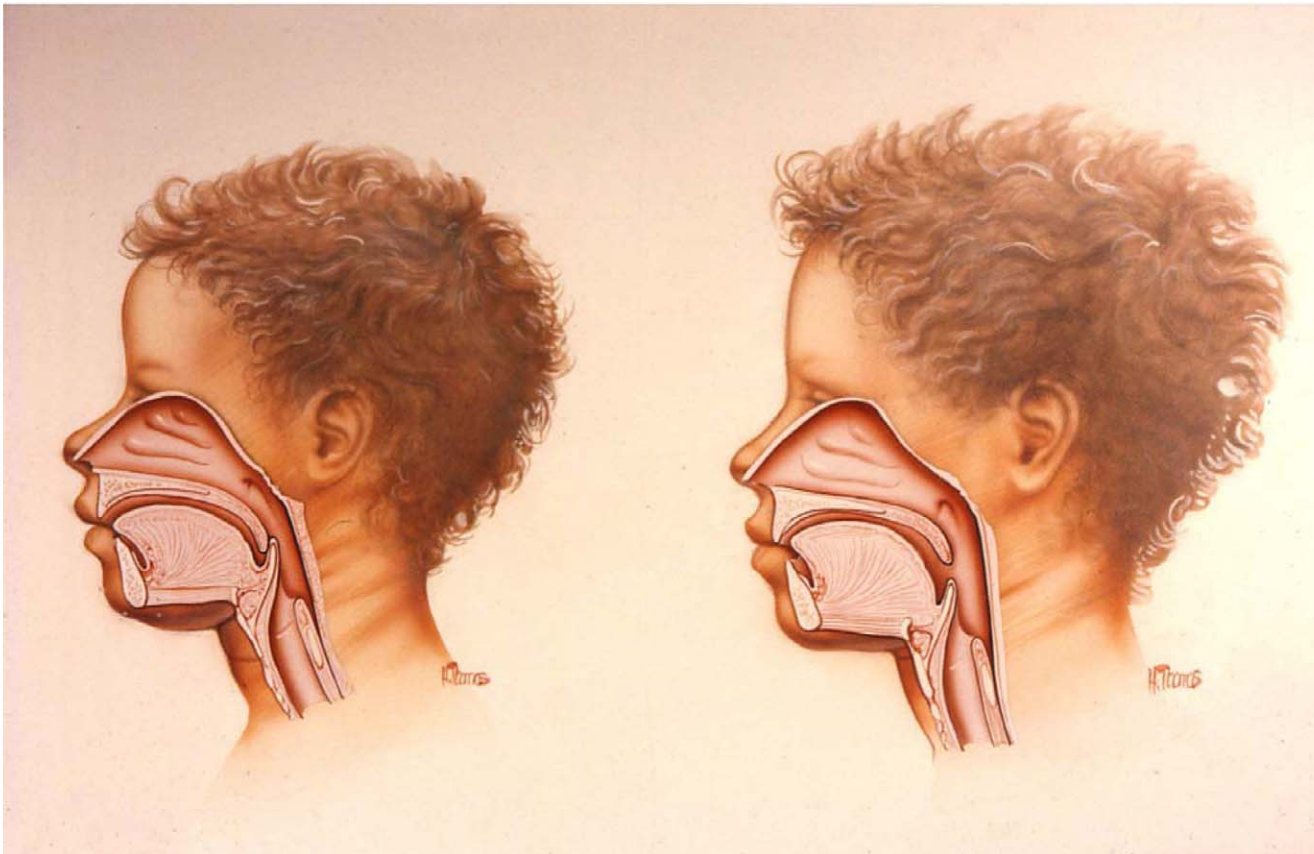

Figure 3. The descent of the larynx.
Source: Image copyright Jeffrey T. Laitman, PhD. Reprinted with permission. The diagram on the left shows the abutting relationship of the soft palate and epiglottis in a child—a configuration common to animal species. On the right, the separation between the soft palate and epiglottis can be seen in an older child.

> newborn human or monkey, would have a more restricted range of vocalizations available than would individuals with larynges placed lower in the neck. (pp. 20-21)

It appears that the quantum vowels [a], [i], and [u], which are the limiting articulations of a vowel triangle that are language-universal, cannot be produced by nonhuman primates or human infants. The absence of these vowels would dramatically restrict the ability to produce speech (Laitman & Reidenberg, 1993; Lieberman, Laitman, Reidenberg, & Gannon, 1992).

The descent of the human larynx is the unique physical characteristic that has allowed the human species to develop speech and, by extension, ultimately to develop language. Similarly, the development of speech and language may have been critical to the creation of song. Although, clearly mental ability plays a substantial role in speech development, the developing human brain would not be able to produce speech without this anatomic distinction. This development may have also played a role in overall human competitiveness and survival. There has been considerable speculation as to the cause of the extinction of *Homo neanderthalensus*. Neanderthals appeared to be larger than humans, walked upright, and had opposable thumbs. Their craniums of 1200 to 1700 cc were larger than those of *Homo sapiens* and 10% greater than modern humans. Most theories support a competitive advantage for *Homo sapiens* that allowed them to propagate and survive (Diamond, 1992; Lahr & Foley, 1998). One prominent theory is that endogenous division of labor and subsequent trading among early modern humans could have helped them to overcome potential biological deficiencies (Horan, Bulte, & Shogren, 2005). This would not seem possible without the ability to communicate with a sophisticated language. The ability to develop speech would lead to the ability to develop language, which may have resulted in a competitive advantage over other species.

It can be argued that the advantages of the ability to breathe and eat simultaneously, the protection from predators, and the prevention of aspiration were far outweighed by evolutionarily benefit of the ability to develop fluent speech. The ability to speak leads to the creation of spoken language, which is necessary to the development of organized civilization. Singing is also a byproduct of the creation of language and was probably the necessary first step in the development of music. Once humans were organized, they could develop culture and the arts (Diamond, 1992). With the exception perhaps of percussion sound, singing was probably the first instrument. It could be argued that without speech, there would not be music and that the human singing voice was necessary to develop music and thereby other musical instrumentation. One unique evolutionary anatomic change, the descent of the human larynx, may have been the critical point in development of human civilization. This, along with the unique layered structure of the human vocal fold, is likely what allowed the creation of song and the development of music.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the authorship and/or publication of this article.

Financial Disclosure/Funding

The author(s) received no financial support for the research and/or authorship of this article.

References

Abitbol, J. (2005). *L'odyssee de la voix.* [The odyssey of the voice]. Paris: Robert Laffont.

Benninger, M. S. (2000). Microdissection or microspot CO2 laser for limited vocal fold benign lesions: A prospective randomized trial. *Laryngoscope*, *110*, 1-17.

Benninger, M. S., Alessi, D., Archer, S., Bastian, R., Ford, C., Koufman, J., ... Woo, P. (1996). Vocal fold scarring: Current concepts and management. *Otolaryngology and Head and Neck Surgery*, *115*, 474-482.

Benninger, M. S., & Jacobson, B. (1995). Vocal nodules, microwebs, and surgery. *Journal of Voice*, *9*, 326-331.

Coulton, R. H. (1994). Physiology of phonation. In M. S. Benninger, B. H. Jacobson, & A. F. Johnson (Eds.), *Vocal arts medicine: The care and prevention of professional voice disorders* (pp. 30-59). New York: Thieme Medical Publishers.

Diamond, J. (1992). *The third chimpanzee: The evolution and future of the human animal.* New York: HarperCollins.

Hartnick, C. J., Rehbar, R., & Prasad, V. (2005). Development and maturation of the pediatric human vocal fold lamina propria. *Laryngoscope*, *115*, 4-15.

Hirano, M. (1974). Morphological structure of the vocal cord as a vibrator and its variations. *Folia Phoniatrica et Logopaedica*, *26*, 89-94.

Hirano, M. (1975). Phonosurgery: Basic and clinical investigations. *Otologia (Fukuoka)*, *21*, 239-442.

Hirano, M., Koike, Y., Hirose, S., & Morio, M. (1973). [Structure of the vocal cord as a vibrator]. *Nippon Jibiinkoka Gakkai Kaiho*, *76*, 1341-1348.

Horan, R. D., Bulte, E., & Shogren, J. F. (2005). How trade saved humanity from biological exclusion: An economic theory of Neanderthal extinction. *Journal of Economic Behavior & Organization*, *58*, 1-29.

Jiang, J. (2006). Physiology of vocie production: How does the voice work? In M. S. Benninger & T. Murry (Eds.), *The performer's voice* (pp. 23-32). San Diego: Plural Publishing.

Lahr, M. M., & Foley, R. A. (1998). Towards a theory of modern human origins: Geography, demography, and diversity in recent human evolution. *Yearbook of Physical Anthropology*, *41*, 137-176.

Laitman, J. T., & Reidenberg, J. S. (1993). Specialization of the human respiratory and upper digestive systems as seen through comparative and developmental anatomy. *Dysphagia*, *8*, 318-325.

Laitman, J. T., Van De Water, T. R., & Noden, D. (1995). Formation of the larynx: From homeobox genes to critical periods. In W. J. Gould, J. S. Rubin, G. Korovin, & R. T. Sataloff (Eds.), *Diagnosis and treatment of voice disorders* (pp. 9-23). New York: Igaku-Shoin Medical Publishers.

Lieberman, P., Laitman, J. T., Reidenberg, J. S., & Gannon, P. J. (1992). The anatomy, physiology, acoustics and perception of speech: Essential elements in analysis of the evolution of human speech. *Journal of Human Evolution*, *23*, 447-467.

Prades, J. M., Dumollard, J. M., Duband, S., Timoshenko, A., Richard, C., Dubois, M. D., ... Peoc'h, M. (2009). Lamina propria of the human vocal fold: Histomorphometric study of collagen fibers. [published online ahead of print October 2009]. *Surgical and Radiologic Anatomy*.

Tucker, H. (1994). Gross and microscopic anatomy of the larynx. In M. S. Benninger, B. H. Jacobson, & A. F. Johnson (Eds.), *Vocal arts medicine: The care and prevention of professional voice disorders* (pp. 11-29). New York: Thieme Medical Publishers.

Van de Berg, J. (1958). Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research*, *1*, 227-244.

Bio

Michael S. Benninger, the chairman of the Head and Neck Institute at The Cleveland Clinic, has treated performing artists for more than 25 years and has published three books related to voice care, including *The Performer's Voice*.